

Opportunity for depth in Chinese eugenics debate

A meeting in Beijing should provide a welcome opportunity to move forward in debates about the ethics and science of eugenics. Western scientists and their hosts should make the most of it.

The publisher HarperCollins and its proprietor Rupert Murdoch have in the past two weeks been humiliatingly exposed in their disposing of the memoirs of the ex-governor of Hong Kong and arch-China-critic Chris Patten. It is ironic that the same publisher was responsible for a book, *Wild Swans* by Jung Chang, which, in its matter-of-fact and compellingly detailed account of the experiences of three generations of women in China, presents as devastating (and widely read) an indictment of that country's recent history as any Westerner could hope to muster. But the book also underlines the circumstances of many Chinese people which have led them to support the communist regime and to continue to do so. It is against that context of widespread poverty and technological backwardness, combined with cultural and even spiritual perceptions of the individual quite different to traditional Western values, that Chinese domestic policies need to be considered.

One policy that has caused much debate is the infant and maternal health law which came into effect in 1995. Introduced in the face of concern about a high "burden of disability", it contains measures that will certainly reinforce the already significant improvement in the — by Western standards — calamitous state of health of large numbers of the population. But it contains articles that refer to the need to consider prevention of marriages, enforced contraception and abortions in the face of "serious hereditary diseases" in adults — in particular, autosomal dominant diseases that cannot be detected prenatally.

To some Chinese, at least, the law appears to represent a wholly positive step. One member of the Chinese Academy of Sciences has defended it thus: "China now has a population of 20 million handicapped....Without effective action, China will have an even larger population with serious hereditary diseases and it will naturally impose a grave social problem as regards their livelihood, social and cultural development as a whole and even the quality

of the whole population."

In contrast, the law is seen by many in the West as an unacceptable application of eugenics. Coming hard on the heels of a recent Chinese track-record of brutal suppression of dissent, it has caused considerable and legitimate heart-searching in the research community. An event initially threatened by such concerns, the 18th International Congress on Genetics organized by the International Genetics Federation (IGF) — a federation of learned societies — is nevertheless taking place in Beijing on 10–15 August this year (see www.igf.org for details and deadlines). In order to meet the concerns of participants about the implications of the new law, and to encourage discussion of the consequences within the Chinese research community, it includes a session on the ethics and science of eugenics. But some Western scientists who would otherwise have attended will stay away given this background. One society — Britain's Genetical Society — is boycotting the meeting, not wishing to be seen as endorsing the law.

That is regrettable. Much needs to be aired in public about differing perceptions of disability between East and West, about the practicalities of screening for genetically complex diseases, about the opportunities of and obstacles to gene therapy, and about the increasingly probable development of germline manipulation — a technology that may well be more enthusiastically embraced in East Asia than in the West, given statements by Asian bioethicists. In short, there is a need for breadth and depth in appreciating the issues, to which this meeting can contribute.

The session on eugenics may turn out to be a fig-leaf rather than an opportunity for genuine discussion. If China wants to develop its reputation as a modern scientific nation, then it has a responsibility to work with the IGF to ensure that this does not happen. It would be a sorry thing to see a much-needed dialogue undermined while, altogether less challengingly, Rupert Murdoch's Star TV spreads into as many Chinese homes as can receive it. □

Gleam in Korean gloom

South Korea's president has taken a bold first step by strengthening the science ministry, but problems abound.

President Kim Dae-Jung has given a moment of cheer to Korean scientists amid the gloom of economic crisis. His move to upgrade the status of the Ministry of Science and Technology (MOST) to a full-blooded ministry with equal status to its powerful neighbours, such as that of trade and industry, is long overdue (see page 112). With stronger teeth, MOST must now achieve better coordination of policy and funding across the various science-related ministries. With prudence, the move should also lead to more money for some of MOST's more innovative funding mechanisms, such as the centre-of-excellence programme of the Korean Science and Engineering Foundation and the new Creative Research Initiative (see *Nature* 391, 625; 1998), both of which provide substantial long-term grants for Korea's best researchers

in universities and government institutes.

But much remains to be done. The universities, of which there are well over a hundred, have a multitude of science and engineering departments in which talented and not-so-talented scientists and engineers languish. The whole university research system, which is administered primarily by the conservatively minded education ministry, needs to be overhauled through processes of research assessment (still rare in Korea) and more targeted funding of the type provided by MOST. The economic crisis may help this process. Several lower-rank universities are said to be on the brink of financial collapse having invested too heavily in new campuses and other facilities. This may encourage the government to focus its research funding more selectively and wisely within the universities. □



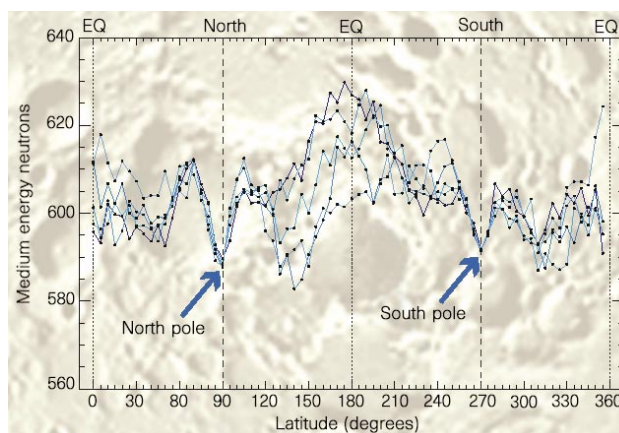
Ice on the Moon boosts hopes for future lunar missions

[WASHINGTON] The apparent discovery announced last week of significant amounts of ice on the Moon may give a boost to several planned lunar missions, including a European concept for a lander funded primarily by private industry. But the announcement by the team responsible for Lunar Prospector appears unlikely to lead government space agencies to invest in lunar bases or other large, expensive undertakings in the near future.

Prospector scientists say they are "extremely certain" they have found between 10 million and 300 million metric tonnes of water mixed in with soil at the lunar poles. But others are urging caution, as the result — based on reduced counts of medium-energy neutrons detected at the poles by the spacecraft's neutron spectrometer — is preliminary, based on only two weeks of data from a mission that will last a year.

Wendell Mendell, a planetary scientist and lunar expert at the NASA Johnson Space Center in Houston, warns that it may be premature to assume the signature is due only to the presence of hydrogen (and therefore water) when the technique is being used on a planetary surface for the first time. Until project scientists gain a better understanding of what other elements on the surface might produce the same effect, he says, "I'm not ready to go start counting ice cubes".

Nevertheless, advocates of lunar exploration say the news is likely to give a lift to international plans for follow-on missions. Japan's Lunar-A spacecraft, for example, scheduled to launch next February, will shoot two penetrators into the Moon's surface to



Telling evidence: early results from Prospector's Neutron Spectrometer show a drop-off in detections of medium-energy neutrons in the polar regions, where hydrogen (indicating water) has absorbed their energy. Project scientists expect the pattern to become more pronounced as more data come in.

study seismology and thermal properties.

It will be followed in 2003 by Selene (Selenological and Engineering Explorer), a joint venture of Japan's two space agencies ISAS and NASDA, which includes a lunar orbiter and lander. Money problems and the failure of the H-2A rocket last month have hampered Japan's space science programme. But the Prospector discovery may improve the missions' chances of survival.

Although Europe has repeatedly expressed an interest in lunar exploration, no mission is currently in the works. Two earlier concepts — the LEDA soft lander and an orbiter called MORO — have been shelved as too expensive. The European Space Agency's (ESA's) new SMART (Small Mission for Advanced Research and Technology) programme, which is meant to fly low-cost technology demonstration spacecraft, may provide the ideal opportunity for a small lunar orbiter

mission. A request for proposals for SMART missions is expected to go out this spring.

More ambitious would be a concept called Euromoon 2000, which has been chosen as a candidate project for a Europe-wide celebration of the millennium (see *Nature* 390, 8; 1997). Although the mission is not supported by space scientists, it has a unique element that boosts its political fortunes — participation by the private sector.

ESA astronaut Wubbo Ockels, who originated the concept in 1996 and now works at the agency's ESTEC technology development centre in the Netherlands, says that three-quarters of the mission's US\$200 million cost will be paid by private sponsors, including aerospace and high-technology companies, as well as Olympics-style corporate sponsors anxious to cash in on the publicity generated by a high-profile return to the Moon.

The concept calls for a small orbiter called Lunarsat to scout south pole landing sites following a launch in 2000. The next year, a lander would be dispatched to a permanently lighted crater rim to begin setting up what Ockels calls a "robotic village" for scientific studies and technology demonstrations. Probes dropped from the lander could also be sent to explore shaded regions where Prospector has detected ice.

Sceptics dismiss the Euromoon concept as a publicity stunt, and say the viability of advertising and television revenues from a space mission is not proven. But Ockels will make his case to the ESA council this month that an investment of around \$8 million for the next phase of the project can keep it on track as it continues to draw more industry partners.

To judge by a press statement from ESA last week, the council may be inclined to grant Ockels the money. On the same day as NASA's announcement, the release outlined European plans for the Moon. At this point, the only plan in sight is Euromoon.

Tony Reichhardt

Fraud claim puts German rules to test

[MUNICH] A laboratory head and one of his technicians resigned last week from the Max Planck Institute for Plant Breeding in Cologne, following accusations of scientific misconduct involving the manipulation of experimental results. The laboratory in question works on the hormonal control of cell division in plants.

The case is the first test of the Max Planck Society's new rules governing the investigation of cases of scientific misconduct. The rules were approved by the society's senate late last year (see *Nature* 390, 430; 1997).

The case came to light after other scientists at the institute became suspicious of the laboratory's apparent success with a particular assay for protoplasts (cultured plant cells without their cell walls) that could

not be repeated by others. They reported their suspicions to the institute's director in February.

Less than three weeks later, the institute has announced publicly — as required by the new rules — that it is embarking on an internal investigation of the laboratory's work. At least one incident of fraud in a published experiment has been confirmed, and the internal inquiry will examine earlier published work.

According to the new rules, the names of the whistleblowers and the accused will not be disclosed during the internal inquiry, which should be completed within four weeks. At that point the institute director will decide whether there is sufficient evidence to ask the society for a formal investigation by independent experts.

Alison Abbott

Insult thwarted 1934 bid to raise profile of Indian science

[NEW DELHI] Ill-tempered remarks by an unnamed British professor at the Indian Institute of Science in Bangalore in the 1930s about the motivation of scientists fleeing Hitler's Germany may have deprived India of a "golden opportunity" to boost its world status as a scientific nation.

The remarks were made in a lecture at a time when Max Born and other scientists were considering emigration to India — then under British rule — to take up faculty positions at the invitation of the late Nobel prizewinner C. V. Raman, who had just become the institute's first Indian director. (Earlier directors were all British.)

Others whom Raman hoped to attract included the physicists Erwin Schrödinger, Rudolf Peierls and Hans Bethe, and the chemist Georg von Hevesy. Born's interest has been recorded by Sivraj Ramaseshan, Raman's nephew and biographer, and editor of *Current Science*, published by the Indian Academy of Sciences. In a recent issue of the journal, Ramaseshan reveals the contents of a letter drafted by Born in reply to Raman's job offer.

In 1934, Raman offered Born a full professorship at the institute. He also invited him to share the directorship of the physics department at a "much higher" salary than Born was receiving as a lecturer at the University of Cambridge. "I have considered [the offer] carefully," Born said in his reply. But he suggested that he "could decide better" if he had "the opportunity of visiting your country, giving lectures for some months".

Raman created an extraordinary visiting chair, and Born went to Bangalore with his wife Hedi. According to Ramaseshan, both soon overcame their doubts about educating their children in India. Born found his students at the Indian Institute of Science intelligent and his lectures were well received. Urged on by Ernest Rutherford, chairman of the institute's selection committee, Born accepted Raman's original offer.

But Raman's delight was to be short lived. According to Ramaseshan, at a meeting organized to welcome the German scientist to the institute, a little known professor referred disparagingly to Born as someone "who was rejected by his own country, a renegade and therefore a second-rate scientist unfit to be part of the faculty, much less to be the head of the department of physics".

"After this public insult Born could not possibly accept Raman's offer," says Ramaseshan. Raman's hopes of getting the other scientists who had expressed an interest in coming to India were also dashed. "India, I feel, missed an incredible golden opportunity," says Ramaseshan. **K. S. Jayaraman**

Korean science ministry gets promised upgrade...

[TOKYO] South Korea's recently elected president, Kim Dae-jung, is living up to a pre-election promise to strengthen the administration of science and technology. Despite political difficulties in establishing his new administration, he has upgraded the status of the Ministry of Science and Technology (MOST), and last week appointed a powerful politician to head the ministry.



Kim: beleaguered but backing science.

Observers of Korean science policy have long argued that MOST lacks the power needed to fulfil its responsibility for coordinating science and technology policy across all ministries. Like Japan's Science and Technology Agency, MOST is more of an 'agency' than a true ministry. It is also surrounded by powerful ministries, such as that of trade and industry, with their own agendas for science and technology.

In 1993 the administration of former president Kim Young-sam talked of restructuring the government and strengthening the administration of science and technology (see *Nature* 364, 377; 1993), but nothing was done. In contrast, even before his inauguration in late February, Kim Dae-jung pushed through the National Assembly a bill that included upgrading MOST to a full ministry.

He even talked of appointing a deputy prime minister of science and technology, alongside the deputies for economic planning and unification. This would have raised MOST to the highest level of government. But in the end the positions of deputy prime minister were removed in the downsizing

and reorganization of government.

In another show of support for science and technology, a few weeks before his inauguration, Kim visited the Korea Institute of Science and Technology (KIST), the country's longest-standing government institute for science and technology, and promised to raise the status of scientists in Korean society.

In line with strengthening MOST, Kim has appointed a powerful politician to head the ministry. The new minister, Kang Chang-hee, is secretary-general of the United Liberal Democrats (ULD) party, the conservative coalition partner of Kim's administration.

Kim made this and other cabinet appointments — despite a deadlock with the majority opposition party about his nomination of his conservative ally, Kim Jong-Pil, honorary president of the ULD, as prime minister — by persuading the outgoing prime minister to propose the new cabinet members.

Some scientists at KIST had been hoping that one of the institute's former presidents might head the new ministry. Nevertheless, one KIST scientist comments that, as a politician, Kang may be a more effective head of the ministry than any scientist. Kang, a graduate and former professor of the Korea Military Academy, headed a national assembly committee on communications and science in 1996, and has been involved in the development of science and technology policy.

With Korea still suffering a severe economic crisis, however, Kang faces an uphill task in strengthening science and technology at a time when government budgets have been cut and their purchasing power for importing scientific equipment and reagents, and for recruiting overseas personnel, have been halved because of the collapse of the won. **David Swinbanks**

...as overstretched universities face bankruptcy

[TOKYO] The economic crisis in South Korea has left several universities facing large debts as they struggle to pay back loans and leases for new facilities and high-tech equipment, according to a report in the Korean media.

The private Dankook University is said to be the first educational institution in Korea to go under, declaring bankruptcy last Saturday (7 March). Dankook was said to have been unable to pay

rising interest rates on huge loans taken out to build a campus on the outskirts of Seoul and a medical school in Chonan in South Chunchong Province.

Dollar-based repayments on loans have risen in recent months by up to 60 per cent, or 100 billion won (\$62.5 million), because of the collapse in the value of the won. A report in the *Korea Herald* speculates that as many as ten universities may be on the verge of collapse.

Universities that have recently established medical schools and leased expensive medical equipment from overseas appear to have been hardest hit. Their problems are compounded by a reduction in subsidies because of cuts in the government's budget required by the bail-out of Korea led by the International Monetary Fund, and the growing numbers of students unable to pay fees in the economic crisis. **D. S.**

French clone provides support for Dolly

[PARIS] Controversy about the authenticity of Dolly, the lamb cloned from an adult udder cell, took a new turn last week with the release by a French group of preliminary data appearing to confirm that differentiated somatic mammalian cells can be reprogrammed to make them totipotent.

In experiments carried out by Jean-Paul Renard's team at the national agricultural research agency (INRA)'s centre at Jouy-en-Josas near Paris, 35 cows were implanted with 61 blastocysts created by somatic cell transfer into enucleated oocytes.

One calf, Marguerite, cloned from a muscle cell of a 60-day-old fetus, has already been born. Four other pregnancies have advanced beyond mid-term, including one obtained from a fetal skin cell containing a transgenic marker, and one from a skin cell from a two-week-old calf.

Renard says the group took the unusual step of releasing results before publication to contribute "supplementary arguments" to the debate triggered by a letter in *Science* by Norton Zinder, a bacterial geneticist at Rockefeller University in New York, and Vittorio Sgaramella, a scientist at the University of Calabria in Italy (see *Nature* 391, 825; 1998).

The letter described Dolly as an "anecdote not a result", given that it was the only successful birth out of several hundred attempts. In particular, the authors criticized what they claimed was the poor characterization of the donor cells used, and the possibility that Dolly could have originated from a stem cell or perhaps even from a circulating fetal cell, as the donor ewe was pregnant when the cells were obtained.

The French experiments, designed to prove that cloning from adult cells is possible, answer many of these criticisms, says Renard. The possibility that the adult clone might be derived from a fetal cell is excluded, for example, because the cells used came from a two-week-old calf that was not pregnant.

Similarly, the donor cells were carefully characterized using marker antibodies against vimentin, which is expressed in fibroblasts, against cytokeratin 3/18, which is expressed only in epithelial cells, and against type A/C lamins, which are expressed only in the nuclei of differentiated cells, not in embryonic cells.

"We can exclude the possibility of any contamination from fetal cells; and the nuclei definitely come from somatic cells, not germinal or fetal ones," says Renard. He adds that the definitive evidence they have cloned from an adult cell must await the birth of the calf, and a genetic fingerprinting comparison with the source animals.

The fact that clones have been obtained from a variety of cell types also suggests that the reprogramming of differentiated soma-



Renard with Marguerite: claims his research answers many of the criticisms of the Dolly scientists.

tic cells to become totipotent is a widespread phenomenon, says Renard. This is consistent with the claims of supporters of Dolly's authenticity, who argue that Dolly is part of a progression from the production of clones from embryos and differentiated fetal cells to the routine generation of embryos from adult somatic cells.

Commenting on the French results, Zinder said last week that the researchers should not count their chickens until they hatch, as

the pregnancies, although advanced, may yet fail. But he also says he has never questioned the possibility that cloning from adult cells may be a reality, and that his concerns focused only on the adequacy of the evidence presented in the Dolly paper.

Renard says he felt uncomfortable about releasing the data before having a paper accepted by a peer-reviewed journal. But he says he eventually felt it necessary to contribute to the controversy. **Declan Butler**

Congressman launches plagiarism inquiry

[WASHINGTON] An investigation by George Brown (Democrat, California), the senior minority member on the House of Representatives Science Committee, into an allegation of scientific plagiarism at Cornell University, New York State, may be widened to look more broadly at cases in which the ideas of junior scientists are allegedly stolen by senior colleagues.

Brown's staff are investigating a case at Cornell in which Antonia Demas, a graduate student, alleges that David Levitsky, a professor of nutrition at the university, failed to give her due credit for research into techniques for improving children's diets. Levitsky vigorously denies the charge.

The staff are also gathering information about other cases, including one at the University of Michigan which last year resulted in the award of \$1.1 million in damages by a jury to Carolyn Phinney, a statistician who had worked in a psychology laboratory, after she claimed that ideas were stolen by her supervising professor.

"It is premature to come to conclusions regarding the Demas case," Brown says. "But the allegations highlight the general problem of young scholars having their best ideas stolen by senior researchers."

Cornell says it has already investigated the case thoroughly. "This set of issues has been repeatedly reviewed by many different

entities," says Henrik Dullea, vice-president for university relations. "The charges were found to be unsupported," he adds, describing it as "ludicrous" to suggest that a university such as Cornell cannot investigate such allegations independently.

"If it is being suggested that I plagiarized or attempted to steal anything, that is totally untrue," says Levitsky. The inspector general at the US Department of Agriculture, which supported the research in question, also investigated the case and took no action.

But Brown believes the affair raises general issues that deserve attention. According to a member of his staff, Brown has no desire to follow in the footsteps of John Dingell (Democrat, Michigan), who fought a wide-ranging battle against what he saw as various forms of scientific misconduct for years. Instead, he will focus on junior researchers — especially women — who may be exploited by senior colleagues.

"There are many reports of such behaviour, but we cannot know how many incidents go unreported due to intimidation or fear," said Brown in a statement. "We cannot allow young scholars to be victimized. Raising questions about cases of alleged misconduct can be good for the integrity of the overall system by providing hope to victims, and giving pause to those who would be predators." **Colin Macilwain**

Prospects brighten for France's Soleil synchrotron facility

[PARIS] French plans to build a new national synchrotron are likely to be approved after next month's regional elections, according to several sources in the synchrotron user community. Plans for the 2.15 GeV Soleil synchrotron, which will cost FF1 billion (US\$ 160 million), were suspended last year by the Socialist government.

Yves Farge, a former ministerial adviser and currently head of research at Pechiney, is resigning to become head of a joint commission between the French Atomic Energy Commission (CEA), the Centre National de la Recherche (CNRS) and the ministry to consider synchrotron radiation in France. Eventually, it is believed, he will head the Soleil project.

Soleil is intended to replace the ageing Lure facilities near Paris, of which Farge was the founder, which has 800 MeV and 1.85 GeV synchrotrons.

The new synchrotron facility has received the backing of Catherine Bréchnac, the director general of CNRS, and Yannick d'Escatha, director general of CEA. But to the dismay of many scientists, the scheme was frozen in November by Claude Allègre, the minister for national education, research and technology, as part of a review of spending on 'big science' (see *Nature* 390, 212; 1997).

Allègre said at the time that large facilities could only be run economically if they were in operation for 24 hours a day, as was the case for international facilities but not national ones. The creation of the latter, he declared would "become the exception". For this reason, he said, the Soleil project would be made "dormant", although not cancelled as such.

But many scientists criticized the decision as being out of touch with the "huge demand" in many research communities. They pointed out that the Soleil project design explicitly specifies that it will operate 24 hours a day, and that synchrotrons are not big equipment used by a few, but a shared tool used by many disciplines.

The underlying reasons for the delay seem to be political, however. Allègre is said to have been keen to avoid taking a decision on the location of the synchrotron — 42 towns have applied to host the facility — until after next month's regional elections.

According to several scientists, a decision to approve Soleil will be announced at that point. Recent statements by Allègre have suggested that his own preference for the location of the machine is in the Paris region, where existing facilities are concentrated.

Declan Butler



Hot water: Southampton Oceanography Centre was completed two years late, and over budget.

NERC in the dock over ocean research centre

[LONDON] Britain's Natural Environment Research Council (NERC) has been accused of "parasitizing" taxpayers during the construction of a £49-million (US\$80-million) oceanography research centre off the south coast of England. The centre was completed nearly two years behind schedule, and was £2.7 million over budget.

At a public hearing last week, members of the House of Commons' Public Accounts Committee described the construction of the Southampton Oceanography Centre as a "sorry tale" of scientists with little business experience losing control of the construction of a multi-million-pound facility.

One committee member, Charles Wardle (Conservative MP for Bexhill and Battle), said that, although he respected the council's scientific expertise, he felt it had used public funds unnecessarily and allowed itself in turn to be "parasitized" by the architects and builders. Another member, Phil Hope (Labour, Corby), castigated council members for "wasting taxpayers' money" and having "their heads in the clouds" when important decisions needed to be taken.

In response, John Krebs, NERC's chief executive, told the committee that the council had learnt its lessons from the affair, and that, since he had taken over in 1994, new guidelines had been adopted and past mistakes would not be repeated. He added that despite the difficult start, the centre had rapidly become a "world class facility".

Andrew Love (Labour, Edmonton) asked how the centre could be described as a "world class" facility when it lacked a proper lecture theatre or auditorium. Krebs responded by saying that a measure of the centre's quality was its ability to attract "superstar" scientists from other countries, such as the United States, despite offering lower salaries.

The committee had summoned NERC officials for questioning following a government audit report last month on the centre's

construction. The facility was designed to provide research and teaching in oceanography, as well as acting as a berth for research vessels.

The report criticized the council for not supervising the project properly. It revealed that the council failed to appoint an overall project manager, despite repeated advice. Sir John Knill, NERC's chairman at the time, jointly managed the project with the vice-chancellor of the University of Southampton.

According to the report, the council allowed builders and architects to make changes to the project without consulting NERC, leading to increased costs. The architects' fees doubled to £8.6 million. The construction company Wimpey claims it is still owed an additional £12.6 million. NERC is challenging the claim.

The report also states that the project had to be scaled down to meet these rising costs. Components of the original design — such as a conference centre and a facility to display the centre's research to the public — had to be omitted, it says, as NERC could not afford to build to the original specification.

But John Woods, who was NERC's director of marine sciences when the centre was constructed, denies this, describing the changes instead as a process of "continuous refinement".

Woods, who is now dean of the graduate school of the environment at Imperial College, London, says that heads of departments were first asked to write down their "dream list" of facilities. "They all knew that they wouldn't get what they wanted. But it was a starting point."

He argues that the conference room and public understanding centre were deliberately omitted as he felt they could be built with private finance. This would leave more of the public funds to be spent on 'scientific' facilities that would be potentially less attractive to a private donor.

Ehsan Masood

French astronomers to rank facilities

[MUNICH] French astronomers are about to carry out a scientific ranking of the ground-based and space-based projects in which they participate — or are considering participation — to help to target required cuts totalling FFr20 million (US\$3.3 million) in this year's astronomy budget.

The move is the first step towards a co-ordinated strategy for astronomy in France and, according to some, could become a model for a Europe-wide effort. It will involve about 200 astronomers, and will take place at a meeting this week organized by the Centre National de la Recherche Scientifique (CNRS) in Arcachon, southwest France.

Last autumn, France's astronomy community was shocked by a demand from the research minister, Claude Allègre, that France cut its annual subscription to the European Southern Observatory (ESO) by a quarter, from FFr130 million to FFr100 million.

But ESO's budget is shared between states according to a fixed formula, and Allègre was persuaded that such a large cut across the board would be rejected. Instead, at an extraordinary meeting of ESO's council last month the representatives of ESO's eight member states agreed to cut 2.5 per cent from each state's contributions (see below).

Allègre was appeased by a suggestion brokered by Jean-François Minster, director of the Institut des Sciences de l'Univers. This involves saving a third of the FFr30 million he demanded by withdrawing French participation from two instruments on ESO's Very Large Telescope (VLT), the building of which is well under way.

The instruments, which are paid for from national rather than ESO budgets, are a



spectrograph and the third auxiliary telescope of the VLT's state-of-the-art interferometry system. France's research ministry must decide in the next few weeks whether to endorse these cuts, and also where to make a further cut of FFr20 million.

The ranking of projects to be conducted at the Arcachon meeting will help in making these immediate decisions. But it will also form a basis for a new strategy for French astronomy in the next decade, with wider implications for European and international joint projects.

Bernard Fort, director of the Institute of Astrophysics of the Observatoire de Paris, and French representative on ESO's council, says the field is wide open. "The meeting will consider the scientific output of big space-based and ground-based planned projects like FIRST [the European Space Agency's planned Far Infrared Observatory], the large millimetre arrays, the next generation space

telescope, and other facilities which all observe the same astronomical events, the formation of stars and galaxies," he says.

One idea being floated is a reduction in French support for instruments on FIRST, which is scheduled to cost France FFr300 million on top of its ESA subscription.

Minster says the meeting will discuss how France can best address important middle-term questions, "and how we must change the structure of French laboratories to achieve this". For example, he says, "we must consider issues such as whether there are too many ground-based telescopes in the international arena, and where to concentrate French efforts".

Given the strain on astronomy budgets throughout Europe, some are now arguing for a Europe-wide ranking exercise of ground-based and space-based astronomy, analogous to the Bahcall report conducted by the National Science Foundation in the United States in the early 1990s.

For example, Steven Beckwith, a director of the Max Planck Institute for Astronomy in Heidelberg, and the newly elected director of the Hubble Space Telescope, says that such a ranking "is long overdue in Europe". ESA's science director, Roger Bonnet, says that ESA's science plan is broadly coordinated with ESO priorities, as well as those of other communities with interest in space science, and so no additional ranking is necessary.

Some astronomers insist that such an exercise should not be conducted across Europe, as each country should decide its own spending according to criteria that are not merely scientific, but may also include industrial interests.

Lodjéwik Woltjer, former chairman of ESA's Space Science Advisory Committee, and an associate of the Observatoire de Haute Provence, says the organization of astronomy in Europe is much more complicated than in the United States. "You couldn't afford to give a negative priority to any project in which one country has particular scientific — and industrial — interests."

But others argue that a European vision should supersede national interests. Having already joined forces with ESA and ESO, European astronomers should "acknowledge European rather than national priorities", says Giovanni Bignami, science director of the Italian Space Agency and professor of astrophysics at the University of Pavia.

But there is a further obstacle to any Europe-wide astronomy ranking, namely the lack of an obvious forum in which it could take place. "Perhaps we should consider an independent body like the European Science Foundation", suggests Manfred Otterbein, director of space science at the German space agency DLR.

Alison Abbott

Observatory vows to keep telescope schedule

[MUNICH] The European Southern Observatory (ESO) must tighten its belt this year to keep construction of its important new project, the Very Large Telescope (VLT), on schedule.

This follows a move by ESO's council to cut 2.5 per cent from the organization's budget for 1998 in response to heavy cuts in France's support for 'big science' projects (see above).

But ESO is optimistic that the shortfall will not disrupt its science programme. "We handled a five per cent cut last year when Germany had budgetary problems," says Norbert König, ESO's

administrative director, "and we will handle the new cut". He adds, however, that ESO would find it difficult to respond to regular cuts every time one of its member states has a budgetary crisis.

Operation of the first of VLT's telescopes is still scheduled for May, with full interferometry expected in 2003. But the possible withdrawal of France from the building of key elements of the interferometry system means that ESO must find new partners if it is to keep to its timetable.

ESO is also looking ahead to its scientific goals after VLT. One decision to be

taken in the next year or so is how far to collaborate with the United States on plans for large-millimetre-array telescopes.

But ESO will also consider its general role in European astronomy, seeking a balance with national astronomy programmes.

It is keeping the debate low-key because of the issue's sensitivity. Opinion is polarized between those who think a larger fraction of national astronomy budgets should be targeted towards ESO, and those who see a need to maintain strong independent national programmes.

A.A.

NIH's way of setting priorities endorsed

[WASHINGTON] An expert panel charged with deciding whether the US National Institutes of Health (NIH) should change the way it sets research priorities heard general praise for the agency's methods last week, and seemed averse to recommending dramatic changes when it reports in July.

Last week's meeting marked the launch of a study by the Institute of Medicine (IOM) into whether the way in which the NIH sets priorities for distributing its \$13.65 billion — and growing — budget should be changed, honed or left alone.

Judging from the tenor of the meeting, the committee is more likely to recommend incremental reforms than a dramatic overhaul. "We're not dealing with something that's broken," says Leon Rosenberg, a professor of molecular biology at Princeton University who chairs the 20-member panel. "We're looking for ways to make a good thing better."

NIH directors who addressed the panel suggested one way that might be achieved: by Congress ceasing to burden the agency with legal requirements to spend money on particular diseases. Such constraints "deform" science, they said.

The \$338,000 IOM study was ordered last year by Congress. It was prompted by the perennial tug of war between the scientific community, which prefers a free hand in deciding how it spends research dollars, and advocates for dozens of diseases, who argue that their causes are underfunded by NIH in a process that excludes the public.

Both sides have support in Congress. Last

week Representative John Porter (Republican, Illinois) took Donna Shalala, the secretary of the Department of Health and Human Services, to task for what he alleged was White House meddling in NIH priority-setting. President Bill Clinton "is setting one disease against another" by seeking in his 1999 budget a five-year, 65 per cent increase in cancer research spending at the NIH, said Porter.

Porter, who chairs the House of Representatives appropriations subcommittee that funds the NIH, said Clinton had opened "a Pandora's box" with the proposal. Shalala defended the proposed increase as "not inappropriate" because, during the same five-year period, overall NIH spending would increase by 48 per cent. She said "cancer is on the cusp of a series of major breakthroughs and this additional investment will make a major difference in the quality of life."

The issue is also coming to a head this year because of the concerns of Senator Bill Frist (Republican, Tennessee), a medical doctor who chairs the subcommittee on public health and safety of the Senate Labor and Human Resources Committee, and who is responsible for drafting a large "reauthorization" bill for the NIH.

The bill will set broad directions for NIH's work over several years. Frist wants to use the IOM study to explain to Congress how NIH reaches its decisions. In general he opposes attempts to direct NIH to spend specific sums on research on specific diseases.

Frist's aide Anne Phelps told the IOM panel that "the whole impetus for the study"

is to have a "process in place" against which to examine proposals for specific disease spending by members of Congress "other than debate on the Senate floor".

Phelps added that, with broad congressional support for doubling the NIH budget over five to ten years, Frist is feeling pressure from disease advocacy groups and their congressional supporters. She expressed a worry that members of Congress will pick five or six areas of research emphasis in the reauthorization bill. "We are very concerned about what will be missing."

There was praise for Harold Varmus, the NIH director, from Adam Yarmolinsky, a panel member who is a professor of public policy at the University of Maryland. He told Varmus at the meeting that he was doing a "superior" job of prioritizing.

Yarmolinsky suggested that NIH does not need the panel's "ideas about how to prioritize better," but rather its advice on how to justify the agency's decisions to the public.

Varmus asked the panel to consider whether there should be ways to make "more radical changes", having effects as big as fusing institutes or halving institutes' budgets. Although such drastic measures are "impossible politically", he said, the panel could consider other ways of achieving change if it did not think the changes NIH makes from year to year are "big enough".

On the issue of NIH responsiveness to the public, Varmus said he has learned "a lot" from highly educated disease advocacy groups. But he added: "There is a point beyond which the decision has to be reserved" for agency officials.

The institute directors who addressed the panel rallied around the current NIH process. Priority setting "is a complicated process, but you want it to be," said Francis Collins, director of the National Human Genome Research Institute. He likened any dramatically simpler scheme to landing a 747 aeroplane with a "joystick".

The directors warned against Congress earmarking research funds. Steven Hyman, director of the National Institute of Mental Health, said that a congressional requirement to devote 15 per cent of research spending to health services research, "lost [us] a substantial percentage of our basic neuroscience portfolio. People simply went away." Earmarking "entrenches self-interested constituencies" and "deforms" science, Hyman added.

But all the directors conceded an important role for public input into the NIH process. John Alderete, a panel member who is a professor of microbiology at the University of Texas Health Science Center at San Antonio, said NIH needs to be more responsive to the public if it wishes to avoid earmarking of funds.

Meredith Wadman

Australia reviews medical research funding

[SYDNEY] The Australian government has responded to intense lobbying against a large cut in funds for medical research, forecast in the May 1997 budget, by setting up the country's first 'strategic' review of health and medical research in 20 years.

The health minister, Michael Wooldridge, last week appointed 13 leading researchers, including three international experts, to an independent panel charged with identifying likely future developments, advising on how "to ensure a continuing research capacity in Australia matched to need" and developing an economic framework to achieve this.

Wooldridge says "there is no magic pudding" for



Wooldridge: no magic, just a review.

research money, but that he was motivated "to take stock" by worries about career prospects for young researchers and the "major squeeze" on medical research.

The panel is chaired by Peter Wills, a businessman who also chairs the Garvan Institute of Medical Research in Sydney. The panel has a budget of A\$1 million (US\$670,000) and will report

by the end of the year. Wills aims "to demonstrate that money put into research is an investment, through, for example, developing better preventive techniques to defray health-care costs".

Peter Doherty, an Australian expatriate and a 1996 Nobel prizewinner, played a key role in the lobbying. Speaking from Memphis, Tennessee, he welcomed the latest move as "putting the subject clearly on the political agenda". Contrasting Australia's approach with recent US funding increases, he says: "The Australian national effort is, at best, in a holding pattern. Basic research and infrastructure are underfunded." **Peter Pockley**

Bid to block Yellowstone enzymes deal

[WASHINGTON] Public interest groups opposed to what they see as “commercialization” of US national parks are taking court action to try to block a deal between Yellowstone National Park and a Californian company that wants to search for and exploit enzymes from the park’s hot springs.

In their lawsuit, filed last week in Washington DC, the groups allege that the Department of the Interior and the National Park Service have breached the 1986 Technology Transfer Act and various other statutes by signing the deal.

The agreement between Yellowstone and Diversa of San Diego, California, was announced by Vice-President Al Gore at Yellowstone’s 125th birthday party last August, and is the first of its type in the United States. It is particularly important because of the vast scale and variety of Yellowstone’s geothermal springs (see right) and the commercial success of the thermostable enzyme *Taq* polymerase. The enzyme is extracted from the bacterium *Thermus aquaticus*, which was found in hot springs at Yellowstone 30 years ago (see *Nature* 364, 2; 1993).

Taq polymerase is now used ubiquitously in the polymerase chain reaction for amplifying DNA. The patents for its use are held by Hoffman LaRoche and, although still subject to legal challenges, might eventually earn the Swiss drug corporation up to \$1 billion a year, none of which will go to Yellowstone.

Beth Burrows, president of the Edmonds Institute, a conservation group based in



Trouble brewing: hot springs conflict is part of fight against national park ‘commercialization’.

Washington state and one of the groups bringing the lawsuit, says: “We are objecting because there was no open, democratic procedure to negotiate the deal, or to decide if Yellowstone should be subject to such deals.”

The lawsuit is also backed by the Alliance of the Wild Rockies, which claims to represent 3,500 users of the Yellowstone National Park, as well as by the International Center for Technology Assessment, which is a group in Washington DC that provides technical and legal support for environmental causes.

Joe Mendelson, a lawyer for the centre, says that Yellowstone is planning half a dozen more agreements of a similar nature, and that up to 20 could soon be negotiated by the national park system. But David Barna, a spokesman for the National Park Service, denies that other deals are being considered.

Barna says Yellowstone and Diversa will not delay the implementation of their agreement because of the lawsuit. “This has been a very open and public process,” he says. “The only thing we’re not making public is the size of the royalty payments. Diversa thinks such disclosure would harm its business interests.”

The deal is a Co-operative Research and Development Agreement (CRADA) permitted under the Technology Transfer Act, which was written to assist public-private collaboration, in part by allowing government agencies to keep commercial data secret.

The public interest groups claim in their lawsuit that the deal is not a CRADA because no laboratory is involved, and natural resources — rather than technology — will be transferred. A congressional member of staff familiar with the law says it does permit the transfer of such resources. But opponents of the deal also claim that the agreement breaks various laws prohibiting commercial activities in national parks.

A spokeswoman for Diversa refused to comment on the court action. According to Mendelson, the district court of the District of Columbia will take several months to rule on the lawsuit.

Colin Macilwain

Plan for \$100m boost to Israeli biotech comes under threat

[JERUSALEM] Concessions made by the Israeli Prime Minister, Benjamin Netanyahu, to ensure parliamentary approval of his budget have clouded the prospects for a US\$100 million venture capital fund that he promised Israel’s biotechnology industry last September.

Yaron Zelekha, who heads the economic staff in the prime minister’s office, says Netanyahu is solidly behind the initiative. But Zelekha admits that the shortfall in the national budget resulting from the deal in parliament is making it difficult to find the money, and that the chances of the fund being established are “50:50”.

According to Netanyahu’s original announcement, about a fifth of the fund’s capital would be provided by the government, the rest coming from private sources. Haim Aviv, who heads the National Israel Biotechnology Committee, an industry forum, says he is “still hopeful” that the fund will be established, despite the tight budget.

Aviv argues that government support for

biotechnology, a fledgling but rapidly growing industry in Israel, is essential to enable it to expand. A survey submitted to the science ministry says the industry’s sales are projected to be \$800 million in 2000, rising to \$1.8 billion in 2003.

The survey, prepared by the Interdisciplinary Center for Technological Analysis and Forecasting at Tel Aviv University, points to a shortage of investment capital as the industry’s most pressing problem. Senior executives in 28 Israeli companies in the fields of pharmaceuticals, diagnostics, agriculture and veterinary medicine said that a shortage of capital is the factor limiting their growth.

“Venture capital is extremely important for this industry, in relatively large sums,” says Eliana Tapuhi, one of the researchers who conducted the survey. Government funds are an important part of the financial resources for these companies, the survey found.

Such funds come largely in the form of grants from the Ministry of Commerce and

Industry. The survey found that grants constitute 20 to 40 per cent of the funding for about one-third of the companies, and 40 to 60 per cent for 16 per cent of companies.

The survey also found an excess of qualified manpower for the industry. Tapuhi says that means there is untapped potential for expansion of biotechnology and related fields in Israel. Industry executives predicted a 105 per cent annual average growth in investment until 2000, falling to 51.7 per cent a year in 2001 to 2003.

“We need a significant injection of capital to realize our potential,” says Aviv, suggesting that the prime minister’s venture capital fund, although not huge, “can serve as a catalyst for the entire process”. Foreign investors will be more willing to invest, he says, if they see that the government is putting money into the industry.

According to Zelekha, a decision on establishing the fund will be made within weeks. If it does not go through this year, “then we’ll do it next year”.

Haim Watzman

DNA tests remove government doubts over tsar's bones

[MUNICH] Following independent DNA analyses by Russian scientists, the government has officially declared that bones found at the site of the execution of the Russian royal family during the revolution do indeed belong to Tsar Nicholas II.

Previous analyses comparing DNA from bone samples with DNA from blood samples of distant relatives, conducted in the United Kingdom and the United States, had strongly suggested that the bones were authentic, but had not been accepted as conclusive in Russia. Russian scientists have now completed analyses of mitochondrial DNA from the bones, and of mitochondrial DNA samples from stored blood of the tsar's nephew, now deceased.

But the Russian Orthodox Church is objecting to the government's decision to bury the bones on 17 July, 80 years after the execution of the tsar and his family by a Bolshevik firing squad. Given the importance of the bones as holy relics, it wants further genetic tests to be conducted to be more certain of their authenticity.

Evgeni Rogae, director of the Laboratory of Molecular Brain Genetics of the Russian Academy of Sciences, and the scientist responsible for the Russian studies, agrees that further tests would be appropriate. Mitochondrial DNA is inherited through the maternal line, he points out, so new tests that can analyse Y-chromosome microsatellites inherited through the paternal line would "help narrow things down even further and put a stop to controversy".

Hawking condemns US cancellation of SSC

[WASHINGTON] President Bill Clinton last week endorsed the US contribution to Europe's Large Hadron Collider after a lecture delivered by the physicist Stephen Hawking at the White House.

Hawking characterized the 1993 cancellation by Congress of the Superconducting Super Collider (SSC) in Texas as "short-sighted". Clinton responded: "This administration opposed the cancellation of the SSC — but we hope that the Swiss project will take up the slack."

Hawking's lecture was the second in a series being delivered at the White House by prominent intellectuals in the run-up to the year 2000. Hawking said there is a 50:50 chance that physicists will find a unified theory in the next 20 years, and also cracked jokes about the US political system.

But he may have more in common with US politics than he knows: the night before the lecture, Hawking hosted a lavish fund-raising

event for the University of Cambridge at the British Embassy in Washington, and his alma mater was repeatedly mentioned in the lecture itself.

Budget blow to Russia's science and schools

[MOSCOW] The Russian State Duma, the lower chamber of parliament, has at last approved the budget for 1998. Science will receive 11.2 billion roubles (US\$1.9 billion), but this is only 2.65 per cent of all budget expenditures, compared to the 4 per cent that is required by law.

Similarly, the total funding for high schools will be 2 per cent of budget expenditures — even though the law demands that this should be at least 3 per cent. And, in practice, funding for science and education are likely to be even lower, as the Duma adopted an amendment to the budget law allowing the cabinet to reduce any approved figure by up to 8 per cent to cover the state's debts. One per cent has already been deducted for this purpose.

Dutch call a halt to calf cloning experiments

[MUNICH] Only ten days after the birth of two genetically identical cloned calves, the Dutch government last week banned the company Pharming from conducting any further experiments involving nuclear transfer. The company says it now plans to carry out further experiments in the United States.

Last year, a law came into force requiring Dutch researchers to reapply for licences for genetic experiments involving animals, including nuclear transfer, and requiring protocols to be approved by an ethical committee. The agriculture minister, Jozijs van Aartsen, announced that a new licence for Pharming's cloning work would not be granted. A newly created national ethics committee wants Pharming first to prove that producing therapeutic proteins from milk results in better drugs than alternative *in vitro* methods of production.

Marine scientists float plan for Euro lobby

[PARIS] National non-governmental associations of marine scientists in Europe are planning to combine their efforts by creating a European Federation of Marine Science Societies, which will lobby public bodies involved in shaping marine science, in particular the European Commission.

The aim, says Jean-François Pavillon, head of the 200-strong Union des Oceanographes de France, is to air the opinions of grassroots marine researchers in policy-making circles, for example in the training of young scientists, the conditions of

use of research fleets, and policies on managing coastal ecosystems. A core group will meet at the Wimereux marine station in Boulogne sur Mer in April to finalize the goals of the new body.

Germany strengthens research ties with India

[NEW DELHI] India and Germany have signed an agreement to boost scientific cooperation through an increased flow of scientists between the two countries. The Project-based Persons' Exchange Programme, which took effect on 1 March, is aimed at researchers in the two countries who are already working on joint science and technology projects, or are planning to develop such projects.

The agreement will allow researchers of one country to work for up to three years in research institutions of the other, with the host government providing grants.

Glaxo cuts HIV drug cost for developing world

[WASHINGTON] The pharmaceutical giant Glaxo Wellcome is to slash the price of an anti-HIV drug to help improve treatment for pregnant women in developing countries, the company announced last week. The move follows the announcement two weeks ago that a short course of the drug AZT in late pregnancy reduced perinatal HIV transmission by 50 per cent in a study in Thailand (see *Nature* 391, 832; 1998).

Glaxo Wellcome will make AZT (Retrovir) available in developing countries at "significantly lower" prices than in the West. The company is also committed to slashing the price of another anti-HIV drug, 3TC (Epivir), now in trials, if it proves effective. A long course of treatment with AZT costs about US\$800 in the West.

US-French deal to seek gene therapy drugs

[WASHINGTON] Maryland-based genomics company Human Genome Sciences (HGS) and Transgène, the biotechnology company based in Strasbourg, France, have announced a 10-year agreement to collaborate on the development of gene therapy drugs.

The companies will together identify novel genes of promise for gene therapy from an HGS database.

The French company will have exclusive rights to license and sublicense up to ten of these genes, and to develop, manufacture and commercialize any resulting gene therapy products worldwide. Alternatively, the companies may choose to co-develop and co-market some or all of the products. The agreement stipulates that HGS will take a 10 per cent interest in Transgène's equity.

Funding and research performance

Sir — You reported the finding that the scientific performance of research institutes supported by the UK Biotechnology and Biological Sciences Research Council during 1981–94 was higher than that of a group of 15 British universities (*Nature* **390**, 12; 1997). But a comparison between the output of universities and other research organizations does not necessarily provide much insight into the factors responsible for observed differences.

We have carried out a study of the effects on research performance of changes in the funding structure of Flemish universities in the 1980s and early 1990s (available on request). In Flanders, as in many countries, universities receive a 'basic allowance' which they are relatively free to spend as they choose. Since the end of the 1970s, this allowance has remained virtually constant.

At the same time, the government has made additional funds available to university research groups on a competitive basis. You have also reported proposals by the Dutch Minister of Education, Culture and Science to transfer DFL500 million (US\$250 million) of university funding to the National Research Council (*Nature* **390**, 9; 1997).

We established for 345 research departments the amount of research funded from the basic allowance, and the support from external funds, expressed in Full Time Equivalents spent on research (FTE-research). We collected bibliometric data and examined trends in publication productivity, defined as the number of

articles per FTE-research, as well as citation rates.

We found that, during the 1980s and early 1990s, externally funded research increased by 7 per cent a year. It became increasingly concentrated in a small number of departments, while the distribution between departments of research funded from the basic allowance remained constant. In 1989, 18 departments — 5% of the total — accounted for 37% of the externally funded research capacity, but only 11% of the 'basic' research capacity.

Departments with a high international impact have profited much more from external funding than groups with a lower standing. The general aim in Flanders, to introduce more competition into the allocation of funds for academic research, has been successful. But overall productivity has remained constant, mainly because the publication output of the departments with the strongest increase in externally funded research decreased significantly, whereas the ratio of junior to senior scientists increased dramatically, from 1.6 to 3.9.

If these trends continue, a situation may emerge in Flanders in which the research base, especially tenured personnel, normally provided by a university out of its own resources, becomes too small for externally funded research activities. Furthermore, there appears to be a limit to the extent to which control over research funds can be transferred from universities

to external funding agencies without a loss of productivity. If this process is pushed too far, the productivity of the system as a whole may decline, as the conditions in which senior scientists have to work — particularly in departments of high international standing — may make it impossible to maintain the potential quality of their research or to train young scientists.

Also, the increasing importance of external competitive funding may lead to an imbalance in the funding system unless universities develop their own somewhat risky internal research policies, relying more than normally upon external funding patterns. This is illustrated by the fact that the faculty of medicine at the Catholic University of Leuven, which has a strong internal research policy, showed the highest impact and productivity of all faculties in the study.

H. F. Moed

*Centre for Science and Technology Studies,
Leiden University, PO Box 9555,
2300 RB Leiden,
The Netherlands
e-mail: moed@cwts.leidenuniv.nl*

M. Luwel

*Ministry of the Flemish Community,
Brussels, Belgium*

J. A. Houben

*H. Van Den Berghe
Catholic University of Leuven,
Leuven, Belgium*

E. Spruyt

*University of Antwerp,
Antwerp, Belgium*

A different view

Sir — Colin Macilwain's News article of 22 January and your leading article of 29 January (*Nature* **391**, 311–312 & 419; 1998) offer one perspective on the US Department of Energy's Academic Strategic Alliances Program (ASAP). As the senior research administrators of the five universities involved, we offer another.

ASAP research will advance and accelerate high-performance computer simulation of physical systems. Researchers at our universities will develop computational algorithms, software and visualization tools for some of the world's most powerful computers and apply them to challenging problems in science and engineering, many of them important in industry. In the course of this work, graduate students and postdocs will be trained in advanced technologies that will have far-reaching applications throughout society.

The awards to our universities derived from a peer-reviewed competition among 22 proposals selected from 48 pre-proposals by many of the country's leading academic institutions. All the ASAP research is unclassified and in well-established academic fields, many of them (including the explosive shocks and turbulence mentioned in your leading article) covered by leading journals such as *Nature*. All our seminars are open to the public and our results will be published in the open literature.

One of the applications of ASAP research will be to the important national goal of Science Based Stockpile Stewardship. This Department of Energy programme, which has strong bipartisan support in the government, will contribute to national security as the United States continues its halt of nuclear testing and so complies with the Comprehensive Test Ban Treaty.

Further details of our ASAP research

can be found on the Web by following the links at <http://www.llnl.gov/asci-alliances/centers.html>

Richard C. Alkire

*(Vice Chancellor for Research)
University of Illinois at Urbana-Champaign*

Richard K. Koehn

*(Vice President for Research)
University of Utah*

Steven E. Koonin*

*(Vice President and Provost)
California Institute of Technology,
Mail Code 206-31,
Pasadena, California 91125, USA
e-mail: stacey@cco.caltech.edu*

Charles Kruger

*(Vice Provost and Dean of
Research and Graduate Policy)
Stanford University*

Robert J. Zimmer

*(Associate Provost for
Research and Education)
University of Chicago*

*Corresponding author

New Zealand reforms weighed in the balance

Sir — Anyone reading your article about the science reforms in New Zealand (*Nature* **391**, 426–427; 1998) could be forgiven for thinking that their main purpose must have been to provide more scientists with more pay, more equipment and jobs for life, simply because science is a good thing. But that was not the purpose at all.

I would like, first, to dispel a major misperception regarding funding. Your correspondent, Peter Pockley, writes of concern that the government “will cut spending on R&D by as much as NZ\$40 million” and that I admit that future prospects for science and technology are not good because, although there will be no reduction in government spending, I stated that “baselines won’t grow”.

But the future of science and technology in New Zealand is of increasing importance; I have never said its prospects are not good. Already built into the funding ‘baseline’ (a technical budgetary term in this context) are increases in spending for each of the next two years. The possible “cuts” referred to by some would at worst amount to a smaller than expected increase, certainly not an actual reduction in funding.

Funding in the first years of the next decade will be more strongly influenced by the outcome of the recently initiated Foresight Project — whose significance your article seriously underestimates — than by the government’s 1996 commitment to increase funding towards a target of 0.8% of gross domestic product. The comments in your leading article (*Nature* **391**, 419; 1998) are relevant and perceptive. Your call for a “new burst of imagination” for New Zealand science is one with which we have already challenged ourselves, in the form of the Foresight Project. This is intended to encourage strategic thinking about the future by all New Zealanders.

My ministry wants to stimulate understanding and awareness of the future we face as a ‘knowledge society’. The impact of globalization, new technologies and demographic, economic and lifestyle trends can be analysed and discussed. The Foresight Project will allow us to think about a range of future scenarios in a way that will help to identify the factors that will determine whether nations, sectors and individuals can prosper. There must be increased focus at all levels on education, research, science and technology. Sectors across the economy and society will be invited to think about the future, and the project will provide a common framework

within which to do this. In the process, the government will use sector-level thinking about the knowledge and capabilities each will require to succeed as one input into setting priorities for publicly funded science and technology.

The real benefits of the reforms to New Zealand science and technology will not be measured in the numbers of dollars spent or scientists employed. They will be measured in the extent to which both sectors and individuals come to recognize that their future well-being depends on science and technology — and spread this view to their children, their politicians and their boards of directors. Watch this space.

Maurice Williamson

(Minister of Research,
Science and Technology)
*Parliament Buildings,
Wellington, New Zealand*

Sir — Sadly, your analysis of New Zealand science is accurate. While agreeing that underfunding of research is a problem, however, I should like to point out a much more insidious change in New Zealand science — the move from “leadership” to “management”.

I had first-hand experience of this change in 1992. The management at my newly created Crown Research Institute (CRI) told the staff that we had to get rid of

the “science culture” and replace it with a “business culture”, something they have relentlessly done over the past six years to the extent that business managers and accountants are now blithely making crucial decisions about research projects which are not subject to meaningful peer review. Pockets of good science do survive, usually centred on high-calibre scientists capable of exerting leadership in a somewhat alien environment.

This disenfranchising of professionals is, of course, a national sport in New Zealand, as witnessed by the enormous problems encountered in a health system run on a market-driven model dreamt up by our Treasury. During my time in the CRIs, the pressure exerted by government to return a “profit” was so extreme that the accountants used to deduct “profit” from my government research grants on the first day of the financial year and return it to the government on the last day.

In my view, if we want to do good science that makes money, the CRIs need to be dramatically revamped. I, and many of my colleagues, believe that the fatal flaw with the CRIs is that they fall between two camps — they are neither research institutes nor commercially viable businesses. Because they are all absolutely dependent on government funding, much of the Foundation for Research, Science and

Technology (FRST) funding is ring-fenced to ensure their survival and not open to truly competitive bidding by the best scientists with the best ideas.

Rather than cloak in pseudo-business-speak the present monopolistic system, could we not simply face up to the fact that we need government-funded research institutes? If we accept this, we could award say one-half of the current FRST funding direct to the CRIs as block grants to maintain their core facilities, and make the other half the subject of a truly competitive grants system to fund scientists whether in universities, private companies, CRIs or elsewhere.

Until we return scientists to centre stage in the New Zealand research scene I fear we shall be the subject of regular and equally depressing analyses by your journal.

R. J. Wilkins

*Department of Biological Sciences,
University of Waikato,
Hamilton, New Zealand
e-mail: d.wilkins@waikato.ac.nz*

Sir — In attempting to form a view on New Zealand science, it must be remembered that New Zealand is an isolated country with a small population and with little tradition of support for research and development (R&D) from the private sector. It also has many good scientists

whose talents need to be tapped in the best possible way. The special circumstances (and problems) required a New Zealand solution and the government and scientific community are to be applauded for having the courage to try a new way.

Is it the best way? I don't think anybody knows the answer, for the jury will be out for several more years. Have mistakes been made? Undoubtedly. The micromanagement of the early days and the undue power accorded to non-scientists to make judgements on operational scientific issues stand out as the obvious ones. But those shortcomings appear to have been recognized and have to be judged against a national requirement for obtaining value for money, and relevance, from the relatively limited funding available for R&D. Is the balance between basic and applied science right? Again, who knows at this point, but the lack of adequate small grants should be addressed.

Is the New Zealand model appropriate to other countries? Given the fact that it has yet to be shown to work in the long term for New Zealand, the safe answer may be no, but small countries with a limited science base would do well to watch the New Zealand experiment, to benefit from its mistakes and hopefully to gain from its successes.

What steps might be taken to increase

the chances of success? Enhancing the overall level of government funding for R&D, increasing the present level of discretionary funding from 10% to 15% or 20% to enable CRIs to do more basic research to underpin their more applied science; providing a better system of funding small 'starter' projects in the universities; and last, but by no means least, encouraging industry to raise its level of support for R&D.

But all that has a familiar ring, and I could just as well be talking about the United Kingdom (or Australia) as New Zealand.

Peter J. Cook

(Past Director,
British Geological Survey)
Research School of Earth Sciences,
Australian National University,
PO Box 4, Canberra, ACT, Australia

Sir — Analysis of science reforms in New Zealand raises serious questions about the future of science in New Zealand, particularly in the light of indications that government commitments to increase science funding will not be met in the 1998–99 budget. A subsequent response by P. M. Hargreaves, president of the Association of Crown Research Institutes (*Nature* 391, 834; 1998), demands a response.

Hargreaves states that 349 new science positions have been established since the Crown Research Institutes (CRIs) were set up, but a breakdown of this figure reveals that it represents only 126 full-time equivalent research and development science positions (CRIs also have a commercial arm). Furthermore, only 21 additional positions for scientists are represented, a 1.6% increase in four years. In contrast, technical staff increased by 7.8% and support staff by 20%. Further analysis of recently released Ministry of Research, Science and Technology figures indicates that these are continuing trends in that the number of scientists declined by almost 4% between 1994 and 1996 while technicians increased by 10% and support staff by 25%.

The New Zealand Association of Scientists is particularly concerned about these marked shifts in employment in government research institutions "from research staff to non-scientists" referred to in the *Nature* article. These are disturbing trends which suggest that the reforms may have spawned a new wave of managerialism and bureaucracy on the New Zealand science scene.

Mike Berridge

(Secretary)
New Zealand Association of Scientists,
PO Box 1874,
Wellington, New Zealand
e-mail: mimrmb@wnmeds.ac.nz

Change comes all too slowly in Albania

Sir — Albania's first democratic government effectively chose to ignore the Albanian Academy of Sciences, but the Socialist party, which took office last year, has been looking at it more critically.

The government appointed a new temporary academy presidium, comprising a president, two vice-presidents and a general secretary, who were given the task of drafting statutes and of organizing elections for a regular presidium. But the members of this temporary presidium worked for the administration of the Hoxha era, and appear to be working undemocratically.

The government said that the new staff of the presidium, together with new directors of the 12 research institutions under the academy's umbrella, should initiate reforms as soon as possible. The most important issues were the means of selecting institute directors and the presidium and of bringing new members into the ageing body of academicians.

The temporary presidium, however, altered the proposed statutes agreed after discussion with scientists at the institutes. The final version, about to be submitted for approval to the president of Albania, says that institute directors should be appointed by the presidium from a list of nominations from the scientific council of each institute. The presidium itself should be elected by an assembly composed of the directors whom the presidium itself appoints, and academicians. It is feared that the presidium will propose to the assembly a two-year postponement of even this procedure.

The slow reorganization process is detrimental for research. Older academy members, some of whom received their scientific titles and degrees by decree in Hoxha's totalitarian regime, cling to their positions and cannot grasp the new paths that science should follow in a democracy.

Researchers will for a long time face not only financial difficulties but also chaotic "reorganizations" by backward minds.

Betim Muco

Seismological Institute, Tirana, Albania
e-mail: betim@sizmo.tirana.al

Addiction, the tobacco industry and *Nature*

Sir — At the end of 1996, *Nature* published my views about behavioural (that is, non-chemical) addictions and the biopsychosocial nature of addiction (*Nature* 384, 18; 1996). These views were originally submitted as an item of

correspondence under the title "Addicted to anything?". After revision, however, it was published under the title "Nicotine, tobacco and addiction".

I thought this title was a little strange — particularly because there was no mention of nicotine, tobacco or smoking in the article itself — but I was pleased just to have had something published in *Nature*. Since that time, however, a number of events have happened that I feel I should share.

Obviously, the publication of my short article automatically led to entry on academic databases all over the world. As a consequence, anyone who types keywords such as "nicotine", "tobacco" or "addiction" into word searches will eventually come across my contribution. On the positive side, I have received what appears to be a record number of reprint requests from academics all over the world wanting to read my thoughts about addiction. In addition, some of those requesting my article were (quite understandably given the title) members of the tobacco industry.

I have also received many telephone calls from the media and legal firms representing the tobacco industry who have done their database word searching and come up with my name (or rather that of "tobacco" and "nicotine"). With regards to the media, I am generally happy to explain my general views on addiction but would be the first to admit I do not consider myself an "expert" on anything concerning nicotine. However, the number of legal firms that have contacted me is not something I have relished.

The feeling I get is that they want to use my research findings to get themselves "off the hook". The general sequence of events is as follows. A legal firm telephones me to say they would like to speak to me face to face about my views on the psychological nature of addiction. I meet them (usually) in their London offices. They tell me they are looking for "scientific advisers" and/or "expert witnesses" to represent their clients (the tobacco industry). I speak to them for about an hour and explain that just because I believe psychological processes to be fundamental in the explanation of all addictions does not excuse the fact that nicotine is physiologically addictive.

Hopefully, with the word "tobacco" in the title of this piece of correspondence, the legal representatives of the tobacco industry will leave me alone!

Mark Griffiths

Psychology Division,
Nottingham Trent University,
Burton Street,
Nottingham NG1 4BU, UK
e-mail: mark.griffiths@ntu.ac.uk

● The title was broadened to reflect an accompanying letter. — Editor, *Nature*

Who should fund basic technology?

The US government sometimes funds research in private companies. But there has been disagreement about the type of research that should receive public money. A consensus may be beginning to emerge.

Lewis M. Branscomb

Scientists and engineers at Calimetrix, a small company in Emeryville, California, working with Energy Conversion Devices and the Polaroid Corporation in the National Storage Industry Consortium, have developed a way of storing data on a CD-ROM disk that holds ten times more information than normal, with five times the access rate. Called 'pit depth modulation', the technique stores 8 to 16 bits of information in each pit pressed into a CD-ROM, instead of the usual two bits stored in a binary system of coding.

The project received cost-shared funding from the Advanced Technology Program (ATP) of the US Department of Commerce's National Institute for Science and Technology. Yet, two years previously, Congressman Bob Walker, while he was chairman of the Republican-controlled House of Representatives Science Committee, criticized ATP as an example of "corporate welfare" and tried to bring an end to the entire programme.

Was this three-dimensional pit storage project an example of basic technological research bearing fruit through a public-private partnership that deserves public support? Or was it an unjustified subsidy of companies that should be expected to recover the costs from their ultimate commercial profits? When is it appropriate for governments of market-economy nations to support research in private companies for the purpose of enhancing their economic performance in commercial markets?

Government officials, legislators, technologists and policy 'wonks' are gathering at the Massachusetts Institute of Technology this week for a National Innovation Summit conference sponsored by the institute and the Council on Competitiveness to discuss the role of government in commercial innovation. Politicians of both parties are looking for common ground that might allow a truce in the policy wars that pitted a sceptical Republican Congress against the commitment of President Bill Clinton and Vice-President Al Gore to expand federal support for public-private partnerships in commercially relevant research. Many signs of a growing consensus have appeared.

Vernon Ehlers (Republican, Michigan), a former university physics professor, heads a congressional task force to draft a new science policy for the United States. The bipartisan Senate science and technology caucus, comprising Senators William Frist, Pete Domenici, Joseph Lieberman and Jay Rockefeller, recently issued proposed consensus policy

principles to guide public investments in research. Lieberman and Domenici, along with Phil Gramm and Jeff Bingaman, introduced a bill with growing bipartisan sponsorship that sets a congressional target of doubling the federal non-defence research budget in ten years. Clinton, speaking last month to the American Association for the Advancement of Science, described "the largest budget increases for NSF [National Science Foundation] and NIH [National Institutes of Health] in history", packaged as a '21st Century Research Fund for America'.

At a time when the US economy seems to be driving ahead with low inflation and high employment for its sixth year without a sign of the usual business-cycle slump, why is so much attention being paid in both parties to public investment in research to bolster US competitiveness?

Two domestic events have contributed to the search for consensus on renewed growth in government research spending. A backlash from the extreme positions of conservatives in the 104th Congress, combined with euphoria about the prospect of a balanced federal budget, have created an opportunity to stabilize a commitment to public investment in research and technology. There are also compelling pressures to sort out the role of government in commercially relevant research arising from global political and trade forces. More open markets, the accelerating privatization of state enterprises, and new World Trade Organization constraints on state subsidies for research and development (R&D) have intensified the debate.

As global competition drives the largest companies to get their products to the marketplace more quickly, to reduce costs and to respond more effectively to customer needs, they are 'downsizing' their research centres and outsourcing more innovation to smaller companies with special skills but more limited resources. As a result, the overall capacity for science-based innovation in the economy is becoming more dependent on access to government-supported basic research in universities and national laboratories. What kinds of research are most needed to support continued commercial innovation?

The struggle to define the government's role in support of research of potential interest to industry is, perhaps, a problem peculiar to the United States. In Europe there are few challenges to the idea that industrial policies are needed, whereas in the United States there is an ideological presumption that the federal government should restrict itself to legislatively authorized missions

(such as defence, health and the environment) and the support of basic academic science. Nevertheless, university researchers in the United Kingdom feared that Margaret Thatcher's government would bring pressure for more immediate economic returns, leading to the massive Foresight exercise to address the correct balance. In the United States the push by Senator Barbara Mikulski of Maryland for more 'strategic' research in the NSF was greeted with grave misgivings.

There have been many attempts to define a category of research that has evident potential for practical uses but is sufficiently speculative that private capital will not provide adequate support for it. Between the 'pure' science of interest to universities and the commercial product technology funded by industry lies most of the research done in the United States. It is in this 'grey area' of technical research — what might be called the 'useful sciences' or 'exploratory technology' — that there is no consensus. Often called 'strategic research' in



NICHOLAS RIGGS/TELEGRAPH COLOUR LIBRARY

Europe, or 'exploratory technology' by the US Department of Defense, the more recent labels by the Thatcher government ('emerging and generic research') and President George Bush's administration ('pre-competitive, generic research') have done little to clarify the debate.

The language used struggles, mostly unsuccessfully, to differentiate the kinds of research government should support from those that can be left to private investment. The NSF and keepers of government R&D statistics in other countries within the Organization for Economic Cooperation and Development still follow the *Frascati Manual*, which recognizes 'basic research', 'applied research' and 'development'.

But this does little to clarify what the role of government should be, as all the confusion is embraced by the highly ambiguous word 'applied'. When members of Congress are asked to support 'applied research', many balk, understandably. They ask whether applied research should be paid for by the private sector, and whether some — perhaps much — applied research when performed by companies at government expense is more than simply 'corporate welfare'? This second question was posed at the National Forum on Harnessing Science and Technology for America's Economic Future by a senior staff member working with Senator John McCain, chairman of the Senate Commerce Committee. It is not an unreasonable view, if by 'applied' research one means finding a narrow solution to an immediate problem faced by a single company.

But most government-funded 'applied' research is called that simply because the researchers (and those who fund them) believe that the work might turn out to be very useful. The fact that research may be useful does not mean it is not 'basic' or academic, nor that market forces will induce companies to pay for it. When the company that did the work can capture only a modest part of the benefits, this useful research has all the characteristics that make 'basic' research appropriate for government funding.

Much of this useful research also has the property that success, if achieved, would benefit a broad range of institutions, including companies, universities and government laboratories. This is the mark of a public good, especially if the achievement cannot be entirely captured by patents, because it involves the creation of new scientific understanding, which cannot be patented.

Will the push to make national research agendas more industry-friendly divert the best researchers to incremental progress on practical problems from industry of little creative challenge, threatening the geometrical progress that has historically driven the progress of science and technology? Or can government investment in exploratory technological research be subject to the same

The answer lies in focusing on basic research that serves intellectual opportunities and public needs in the most creative ways



standards of creativity we expect from science? I have proposed the label 'basic technology research' to try to capture this picture of the most advanced technology research of potential interest to industry.

Richard Nelson has observed that the legacy of Vannevar Bush in the United States is an abiding suspicion of government funding for research in support of commercial innovation. Outside the province of the military, which has been cutting back on its sponsorship of basic technological research, basic research funding from the government is more likely to support analytical than synthetic chemistry, and research on polymer chemistry rather than polymeric materials. Support for basic chemical engineering research is less than 1% of the federal basic research budget, yet this is one of the most useful fields for the support of industrial productivity. These historical trends reflect the greater interest of the military in product performance than in process technology.

The NIH is expected to receive a whopping 8% increase in the 1999 budget. But why are politicians so enthusiastic about funding biomedical research? Its appeal lies not only in the fact that it addresses diseases to which congressmen are prone (as George Brown said with tongue in cheek), but also in the fact that investments in basic research seem to lead automatically to cures for disease. The institutional integration of clinical research with basic biomedical science can be traced back to the Flexner report on US medical education some 80 years ago. So the highly visible discoveries in basic biomedical science — such as the structure of DNA and the workings of the immune system — are seen by the public as leading more or less automatically to monoclonal antibodies for diagnosis and treatments for cancer.

People view the deliverers of the benefits of biomedical science — the research physicians and nurses — as part of the same research system as the biologists. The public appreciates government investments that lead to dramatic discoveries, especially when the discoveries seem to be linked directly to

equally creative, useful research that brings more immediate benefits in their own lives.

Unfortunately this association of new science with its benefits is not quite as readily accepted in the physical sciences. Although physical sciences are publicly supported, the technologies embodied in products and services are usually delivered by commercial companies financed by private capital. The public is fascinated by fundamental or pure science, especially when it addresses human or cosmic origins and fates. But the public does not recognize the role of engineers and scientists in basic technology research as the equivalent of that of clinical researchers in medicine, and ascribes the benefits of physical science — the computers or jet aircraft — to the companies that make them.

What the public and the politicians fail to understand is that the basic technology research that bridges conceptual progress in science to the required advances in products and services also requires the kind of imaginative research that we associate with the idea of 'basic' research, whether performed by government or industry.

The parsing of government-funded research into 'basic' and 'applied' research and 'development' will be difficult to change. Accountants preparing budgets for Congress are not the only ones interested in preserving these questionable distinctions. Academic scientists and engineers identify their work as 'basic' rather than 'applied' to protect their freedom to explore nature in their own way. The public is often told that many years must elapse before the fruits of physical science will be available.

The scientific community reinforces this impression, because scientists want those providing the funds to be very patient and not to 'micro-manage' science in the quest for outcomes of immediate value, and few scientists want their work described as 'applied'. There is plenty of history to justify the fear of micro-management of research by federal programme managers. These fears need to be respected. Scientists and engineers must recognize that their fears of micro-management are a vital factor in preventing a wider recognition that private investment fails to provide adequate support for basic technological research.

So when is it appropriate for governments to support research in companies? The answer lies in focusing all government science and technology activities (other than those for the military) on basic research — both technological and scientific — that serves both intellectual opportunities and public needs in the most creative ways. This policy should apply not only to universities, but also to public-private partnerships with consortia of companies and universities. □

Lewis M. Branscomb is at the John F. Kennedy School of Government, Harvard University, 79 John F. Kennedy Street, Cambridge, MA 02138, USA.

Tailor-made condensates

Keith Burnett

Bose–Einstein condensates are a strange form of ultracold matter, first created in the lab less than three years ago. A way has now been found to manipulate the interactions within condensates to an extraordinary degree, heralding previously impossible experiments and applications.

One of the favourite games of the theoretical physicist is to imagine that there is a little knob with which one can change the nature of the interactions between the atoms of a material. This is usually just a game, although it is a very instructive one: one sees very clearly in such an exercise how important the interaction strength is for the properties of real materials. For that reason it is something one would dearly like to be able to do in real life.

In new experiments at the Massachusetts Institute of Technology, reported on page 151 of this issue¹, Wolfgang Ketterle's group has shown that, in Bose–Einstein condensates, the strength of the interactions between atoms can indeed be changed over a very wide range, and can even be switched from attractive to repulsive. So for ultracold trapped atoms, the theorists' dream has become reality. This is especially exciting because of the range of phenomena that can be studied as a consequence.

A Bose–Einstein condensed gas of atoms was first produced in 1995 by Eric Cornell, Carl Wieman and colleagues² at JILA in the University of Colorado. They used evaporative cooling of atoms in a magnetic trap to achieve it³. In a Bose–Einstein condensate,

all the atoms share the same quantum state, and it therefore has quite remarkable properties. In particular, it is subject to macroscopic quantum effects like those seen in superfluids, such as persistent viscosity-free flow. The production of such states was built on the development of laser cooling, for which the 1997 Nobel prize was awarded to Claude Cohen-Tannoudji, Steve Chu and Bill Phillips, and it is a remarkable technical achievement. But the properties of condensates are also of great interest to theorists, especially as atomic Bose condensates are so weakly interacting that first-principles microscopic calculations can be applied to their description.

The theoretical possibility of tuning the interaction between ultracold atoms using a

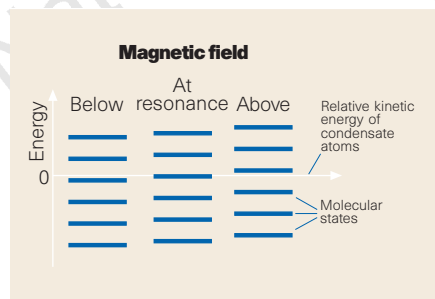


Figure 1 The energies of nearly-bound molecular states for different values of an applied magnetic field. These are states in which the atoms of a Bose–Einstein condensate hang together for some time instead of simply cruising past each other. Two atoms in the condensate are much more likely to form such a pair if their relative kinetic energy is close to the energy of the molecular state — in other words, if there is a resonance. Because a magnetic field alters the molecular-state energies, the strength of atom–atom interactions can be changed enormously by varying the field.

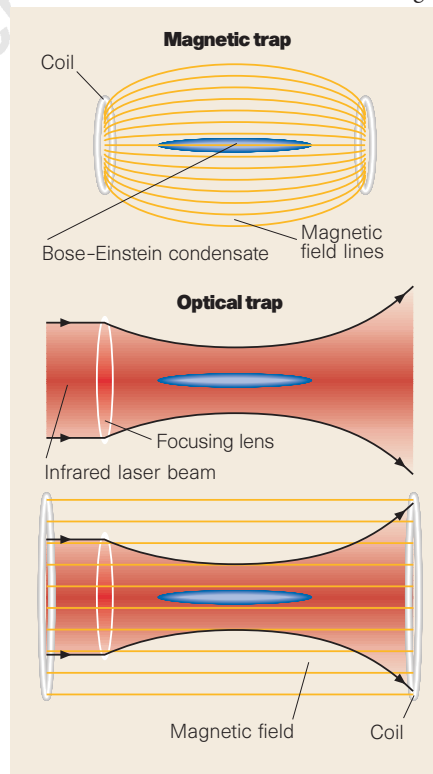


Figure 2 Trapping and tweaking a Bose–Einstein condensate. The condensate is first formed by evaporative and laser cooling in a magnetic trap (a). It is then contained by the optical field of a focused infrared laser (b), so that its interactions can be tuned by a uniform magnetic field (c).

magnetic field was established some time ago³. A modest magnetic field can distort the electron cloud around the atoms and hence change the force they exert on one another. This change is very small, but at the ultracold temperatures of the condensate it has a profound effect (Fig. 1).

To achieve this in the lab required the ability to trap the atoms so that they would not be too badly disturbed by a magnetic tuning field — the usual magnetic trapping methods clearly won't do. Instead, Ketterle *et al.* managed to trap their condensate in a light field. That isn't easy, as light fields easily destroy condensates unless the fields are closely controlled, and its accomplishment is a masterpiece of experimental technique.

Once they had made the condensate in a magnetic trap, the authors transferred it into an optical one, where the atoms are confined solely by the force a laser field exerts on an atom (Fig. 2). It is only possible to use such a trap because of the remarkably low temperatures, below 1 μK , achieved in the first step of creating the condensate. This makes the atoms slow enough for them to be held in a laser beam of modest power.

Ketterle and colleagues could then apply a near-uniform magnetic field (which does not unduly disturb the confinement of the atoms) and observe the changes this produced in the atomic interactions by the effect it has on the properties of the condensate. The changes they see are very marked, and the confirmation of theoretical predictions is impressive.

Being able to tune the interactions offers new ways to control the motion of condensed gases when they are allowed to escape to produce a laser-like source of atoms. By using a magnetic field with a varying field strength, one could focus the matter wave in a clean, loss-free manner. This should lead to more flexible sources of coherent matter waves for clocks, interferometry and lithography, for example.

This tuning knob will also let experimentalists play the game that theorists have been so very fond of in the past. One of the most exciting possibilities would be to form a large, stable condensate with repulsive interactions, and then to switch the interactions to being attractive. That would produce an unstable macroscopic quantum system that would rapidly collapse, ejecting hot atoms and molecules — providing a stringent test of the quantum field theory that is now being developed to treat condensates and their dynamics. □

Keith Burnett is in the Clarendon Laboratory, University of Oxford, Parks Road, Oxford OX1 3PU, UK.

e-mail: k.burnett1@physics.oxford.ac.uk

1. Inoué, S. *et al.* *Nature* **392**, 151–154 (1998).
2. Anderson, M. H., Ensher, J. R., Matthews, M. R., Wieman, C. E. & Cornell, E. A. *Science* **269**, 198–201 (1995).
3. Davis, K. B. *et al.* *Phys. Rev. Lett.* **75**, 3969–3973 (1995).

Human pheromones

Communication through body odour

Aron Weller

Human communication is dominated by auditory and visual information. In contrast, many animals use smell to communicate — both immediate and long-term effects of chemical signals have been documented within many species, from yeasts to mammals. This contrast raises two questions. Are humans well-equipped for broadcasting and receiving social chemical messages? And do we communicate through such messages? A study reported by Stern and McClintock¹ on page 177 of this issue suggests that the answer to both of these questions is 'yes'.

Almost 40 years ago, Karlson and Luscher² coined the term pheromone, meaning a chemical that, when emitted from one animal, exerts a behavioural or physiological response in another animal of the same species. Releaser pheromones stimulate rapid behavioural changes, whereas primer pheromones result, through the neuroendocrine system, in delayed physiological or behavioural changes. Other, information pheromones are indicators of an animal's identity or territory.

Mammals usually (although not exclusively) detect pheromones through receptors found in a specialized structure called the vomeronasal organ (VNO). This is a small tubular structure lined with receptor cells, and it is close to the nasal cavity. Pheromonal information sensed by the VNO is transferred to the accessory olfactory bulb and other regions of the brain, including the anterior part of the hypothalamus. This region controls the neuroendocrine systems responsible for aspects of reproductive physiology and behaviour^{3,4}. The VNO-to-brain pathway constitutes the accessory olfactory system, and it is distinct from the main olfactory system, the receptors for which are in the olfactory epithelium in the nose.

Until a few years ago, the human VNO was considered to be atrophied in adults. However, a clearly identifiable VNO was then found near the base of the nasal septum in adults⁴. Initial studies applying chemicals derived from adult human skin to the VNO showed changes in the autonomic nervous system, and in the periodicity of follicle-stimulating hormone and luteinizing hormone from the pituitary gland^{4,5}. These results indicate that a potentially functional VNO–hypothalamic–pituitary–gonadal axis exists. But can humans actually use this system to process and respond to chemical signals emitted by other humans?

Humans, like other animals, emit odours

from many parts of their bodies⁶. Personal body odour represents secretions from several types of skin gland, most of which are concentrated in the underarm (axillary) area⁷. The biochemical composition of these secretions (and the resulting individual body odour) depends on genetic, hormonal, metabolic, dietary, psychological, social and environmental influences⁶. It is not known whether the olfactory signals from an individual's secretions are perceived consciously, and processed through the main olfactory system, or whether a portion of the signals are pheromones, which are presumably processed unconsciously through the accessory olfactory system.

Studies have shown⁸ that a mother can identify the odour of her newborn infant or older child by smelling a T-shirt worn previously by the child, correctly discriminating it from a shirt worn by another child of the same age. In turn, infants prefer breast or axillary pads from their own mothers over pads from unfamiliar mothers. Moreover, children discriminate and prefer their mother's odour, and the body odours of other kin (such as siblings and grandchildren) can also be differentiated⁶. Thus, body odours can provide important identifying information in humans. But can they serve as releaser or primer pheromones? More specifically, can chemical signals from one human be detected by another without being consciously experienced as an odour? And might

they have immediate, or delayed, effects on the neuroendocrinological reproductive systems of other humans?

Research pioneered by McClintock almost 30 years ago⁹ showed that the menstrual cycles of women who are room-mates or close friends tend to converge over time. This suggests that some factor related to social closeness and interaction can shift the timing of the biological clock in the brain that determines ovulation and cycling. The function of this phenomenon — menstrual synchrony¹⁰ — is unclear¹¹. It has often been quoted as evidence for (primer) pheromonal communication in humans. However, this conclusion is incorrect because the mode of communication between women that results in synchrony is unknown. Although VNO olfaction has been proposed as the most likely candidate, it is also possible that menstrual synchrony is mediated by visual or auditory signals, through mutual social activities, similar daily schedules or exposure to similar stimuli.

Since her initial report⁹, McClintock has been examining reproductive behaviour in rodents. By isolating female rats in cages connected only by a common air supply, she showed¹² that the donor rat produces one signal that accelerates ovulation (and shortens oestrus) in the recipient, and another signal that has the opposite effects. Aided by a computer simulation, Schank and McClintock¹³ proposed a model of how the two signals interact to produce oestrous synchrony. The model spans three levels of organization — the group, the rat and the neuroendocrine component — and relies on two hypothetical pheromones, one that delays and one that advances the phase of the ovarian system. This model is in accordance with other diverse biological systems that show syn-

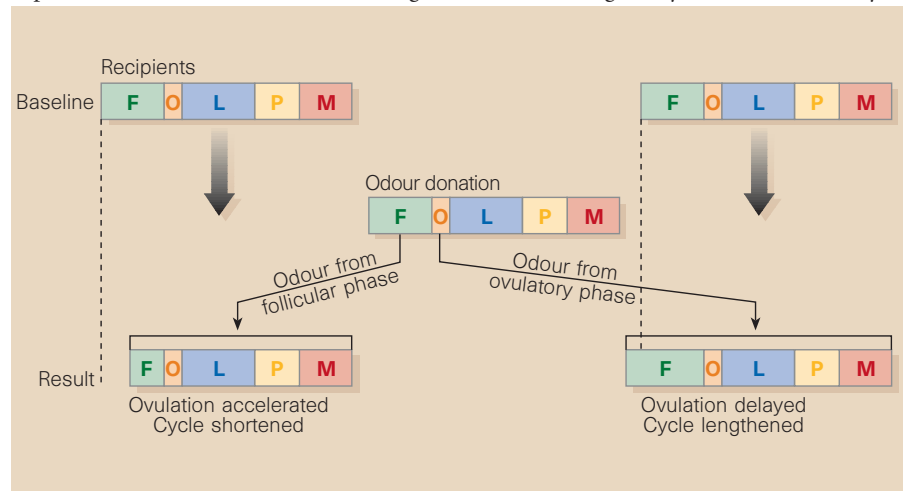


Figure 1 Phases of the menstrual cycle for the women participating in Stern and McClintock's study¹. The menstrual cycle, which is broadly divided into the follicular and luteal phases, is represented here by its sub-phases: follicular (F), ovulatory (O), luteal (L), premenstrual (P) and menstrual (M). Axillary (armpit) odours from the donors' F phase accelerated the latency to ovulation and shortened the cycle length of the recipients (left side). Donations from the O phase delayed ovulation and lengthened the cycles of the recipients (right side). Only the F phase of the recipients was regulated — shortened or lengthened, respectively.

chronization of cycles, such as the flashing of fireflies.

Stern and McClintock¹ now provide support for the generality of the previous model. They collected body odour on cotton pads from female donors. The pads were wiped above the upper lips (under the noses) of other women (recipients), who were asked not to wash their faces for the next six hours. This procedure was repeated daily over two continuous menstrual cycles. The authors found that the biological clocks of the recipients were affected — the timing of their ovulation and menstruation were systematically changed. Specifically, axillary odours from women in the follicular phase of the ovulatory cycle shortened both the time to ovulation and the length of the menstrual cycle in the recipients. Odours taken on the day that the donors ovulated (and the next two days) delayed ovulation and lengthened the total cycle of the recipients (Fig. 1). These phase-advancing and phase-delaying effects show that human axillary compounds can regulate biological rhythms.

This carefully controlled study clearly shows, for the first time, that the potential for chemical communication involving sexual function has been preserved in humans during evolution. Moreover, humans respond to body-odour signals in a neuroendocrinological manner that is similar to (and, in fact, was predicted from) animal models. However, we still do not have evidence that humans actually do communicate by pheromones in modern society. To test this, we could examine whether a phenomenon such as menstrual synchrony exists in women with no sense of smell. Or we could prevent a social odour from acting in a natural social human situation, and assess the result. Elimination of the phenomenon in such a study would support chemical communication. But a negative result would still not exclude the possibility that pheromones are one of the many modes of human communication.

The finding that humans can communicate by pheromones¹ is, nevertheless, ground-breaking and opens many possibilities for future study and application. The active components of body odours (when clearly identified) could be used as natural, alternative ways to control the time of ovulation; for example, as an aid in contraception. Furthermore, as implied by the authors, we may yet discover that other aspects of our behaviour and physiology are affected by covert olfactory messages from other people during social interactions. Because odours have well-known influences on emotions, perhaps human pheromones complement other sources of interpersonal information. This may result in feelings such as emotional contagion¹⁴ (experiencing another person's feelings), sympathy, empathy and their accompanying physiological reactions. □

Aron Weller is in the Department of Psychology, Bar-Ilan University, Ramat-Gan 52900, Israel. e-mail: weller@pop.ey.ac.il

1. Stern, K. & McClintock, M. K. *Nature* **392**, 177–179 (1998).
2. Karlson, P. & Luscher, M. *Nature* **183**, 55–56 (1959).
3. Bartoshuk, L. M. & Beauchamp, G. K. *Annu. Rev. Psychol.* **45**, 419–449 (1994).
4. Monti-Bloch, L., Jennings-White, C., Dolberg, D. S. & Berliner, D. L. *Psychoneuroendocrinology* **19**, 673–686 (1994).
5. Berliner, D. L., Monti-Bloch, L., Jennings-White, C. & Diaz-Sanchez, V. J. *Steroid Biochem. Mol. Biol.* **58**, 259–265 (1996).
6. Schaal, B. & Porter, R. H. in *Advances in the Study of Behavior* 20 (eds Slater, P. J. B., Rosenblatt, J. S., Beer, C. & Milinski, M.) 135–199 (Academic, New York, 1991).

7. Spielman, A. I., Zeng, X.-N., Leyden, J. J. & Preti, G. *Experientia* **51**, 40–47 (1995).
8. Scaal, B. et al. *Reprod. Nutr. Dev.* **20**, 843–858 (1980).
9. McClintock, M. K. *Nature* **229**, 244–245 (1971).
10. Weller, L. & Weller, A. *Neurosci. Biobehav. Rev.* **17**, 427–439 (1993).
11. Weller, A. & Weller, L. J. *Comp. Psychol.* **111**, 143–151 (1997).
12. McClintock, M. K. in *Pheromones and Reproduction in Mammals* (ed. Vandenbergh, J. G.) 113–149 (Academic, New York, 1983).
13. Schank, J. & McClintock, M. K. *J. Theor. Biol.* **157**, 317–362 (1992).
14. Hatfield, E., Cacioppo, J. T. & Rapson, R. L. *Rev. Pers. Soc. Psychol.* **14**, 151–177 (1992).

Condensed matter

One substance, two liquids?

Pablo G. Debenedetti

Water, like any other liquid, can be supercooled — cooled below its freezing point without crystallizing. The physical properties of supercooled water are unusual: the lower its temperature, the easier it is to compress, and the more pronounced its anomalous tendency to expand when cooled. Most other liquids contract when cooled, and are more difficult to compress the lower their temperature. As if these characteristics were not peculiar enough, on page 164 of this issue¹, Mishima and Stanley offer new evidence for the notion that two different forms of supercooled water may coexist. Although coexistence of liquid mixtures with different compositions is common, a phase transition between two liquid forms of a pure substance, both lacking long-range order, has never been observed.

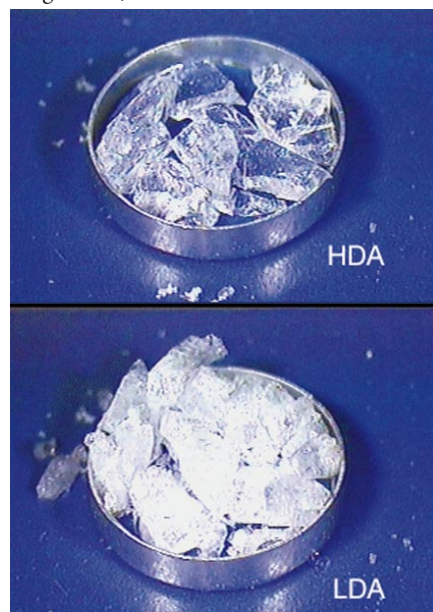


Figure 1 The two forms of glassy water. LDA is formed by rapidly cooling water at atmospheric pressure; HDA is formed by compressing either LDA or ordinary ice at low temperature. These amorphous solids might be able to coexist, as might two structurally similar liquid waters, LDL and HDL.

When liquid water is cooled fast enough to avoid crystallization, it forms a glass². Such vitreous water is found as frost on interstellar dust in dense molecular clouds, and comets are made of it³. In 1985, Mishima and co-workers⁴ proposed the existence of a transition between two forms of glassy water: low-density and high-density amorphous ice (LDA and HDA; Fig. 1). They observed

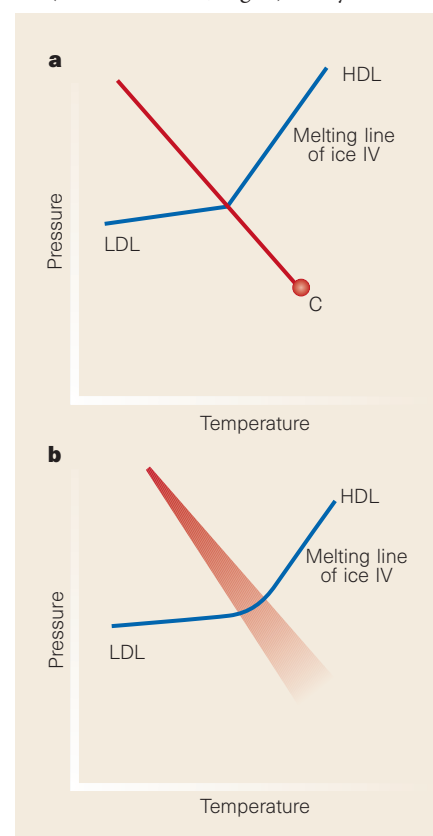


Figure 2 The melting line of ice IV. The sharp change in its slope implies a sudden change in the density and entropy of liquid water¹. This could occur, (a) along a line of coexistence between a low-density and a high-density liquid (LDL, HDL), terminating at a critical point⁵, C; or (b) over a limited range of temperatures and pressures that becomes progressively narrower at low temperatures and high pressures⁷.

sharp changes in volume as a function of pressure, which they interpreted as a phase transition with discontinuities in density and entropy — a first-order transition.

This coexistence between two disordered phases was a bold new idea. In 1992 it was used to explain⁵ certain anomalies of supercooled water⁶, especially the large increase in properties such as specific heat at low temperature. In this interpretation, the coexistence of LDA and HDA is the low-temperature manifestation of an underlying transition between two liquid phases, LDL and HDL. (In each case, the amorphous ice and the liquid would be separated by a glass transition.) The transition between the two liquids would terminate at a critical point (Fig. 2a), where local fluctuations grow in size, causing the thermodynamic response functions to go to infinity. This second critical point of water would be metastable with respect to crystallization⁵.

This hypothesis has remained unproved by experiment, because supercooled water can only be studied above its homogeneous nucleation temperature (-42°C at one atmosphere), at which spontaneous crystallization occurs. To avoid crystallization, it is necessary to cool water very fast² (at about $10^6^{\circ}\text{C s}^{-1}$), and the result is a glass. Mishima and Stanley¹ come very close to overcoming this obstacle. They measure the melting curve of ice IV, a high-pressure form of ice, by decompressing an ice emulsion. The melting curve undergoes a sharp change in slope at the proposed line of liquid–liquid transitions — exactly what should happen if ice IV is melting into two different liquids in different regimes.

Is this proof of the liquid–liquid transition hypothesis? Not quite. The density and entropy of the liquid phase might change abruptly, but not discontinuously (Fig. 2b), over a region whose width decreases with decreasing temperature and increasing pressure⁷. This would also explain the sudden change in the slope of the melting curve of ice IV. A third explanation of supercooled water's anomalies is the stability-limit conjecture⁸, which does not appear to fit these results. In this interpretation, there is a line along which thermodynamic quantities such as compressibility and heat capacity go to infinity, and beyond which liquid water cannot exist.

Discriminating between the two explanations shown in Fig. 2 requires proof or disproof of coexistence between two phases. This is a trivial matter for, say, a vapour and a liquid. But at much lower temperatures, where non-crystalline water exists in vitreous form, the extraordinarily long equilibration times characteristic of glasses make direct proof of coexistence between LDA and HDA problematic. Furthermore, the sudden change in slope of the melting curve of ice IV occurs near to the line of

homogeneous nucleation, where the liquid formed by decompression of ice IV is expected to freeze rapidly. This may call for additional subtleties of interpretation.

Nevertheless, Mishima and Stanley's work provides strong support for the notion that non-crystalline water can exist in two forms. These forms have very different densities: at 77 K, the compression of LDA into HDA occurs with an abrupt decrease in volume⁴ of $\sim 22\%$. Perhaps another parameter is needed to describe the different phases⁹, one that can distinguish differences in local structure.

That the transformation from one form of water to the other is discontinuous — that the two forms can coexist — still remains to be proved. But the models shown in Fig. 2 would have the same microscopic cause: the transient and localized formation of low-density and low-energy structured molecular arrangements stabilized by strong, orientation-dependent interactions such as hydrogen bonding^{7,9,10}.

A definitive physical picture that explains the behaviour of non-crystalline water at low temperatures is still lacking¹¹. This problem is intellectually challenging, because of the

novelty of concepts such as liquid–liquid coexistence in a pure substance; and it is also of great practical significance, because of the ubiquity of supercooled and glassy water in stratospheric clouds and interstellar dust, respectively, and the technological potential of methods for storing labile biochemicals in supercooled water emulsions or in water–carbohydrate glasses. Few questions about the liquid state of matter appear more intriguing or important. □

Pablo G. Debenedetti is in the Chemical Engineering Department, Princeton University, Princeton, New Jersey 08544-5263, USA.
e-mail: pdebene@pucc.princeton.edu

1. Mishima, O. & Stanley, H. E. *Nature* **392**, 164–168 (1998).
2. Bruggeller, P. & Mayer, E. *Nature* **288**, 569–571 (1980).
3. Jenniskens, P. & Blake, D. F. *Science* **265**, 753–756 (1994).
4. Mishima, O., Calvert, L. D. & Whalley, E. *Nature* **314**, 76–78 (1985).
5. Poole, P. H., Sciortino, F., Essman, U. & Stanley, H. E. *Nature* **360**, 324–328 (1992).
6. Speedy, R. J. & Angell, C. A. *J. Chem. Phys.* **65**, 851–858 (1976).
7. Sastry, S., Debenedetti, P. G., Sciortino, F. & Stanley, H. E. *Phys. Rev. E* **53**, 6144–6154 (1996).
8. Speedy, R. J. *J. Phys. Chem.* **86**, 982–991 (1982).
9. Shiratani, E. & Sasaki, M. *J. Chem. Phys.* **108**, 3264–3276 (1998).
10. Poole, P. H., Sciortino, F., Grande, T., Stanley, H. E. & Angell, C. A. *Phys. Rev. Lett.* **73**, 1632–1635 (1994).
11. Johari, G. P., Hallbrucker, A. & Mayer, E. *Science* **273**, 90–92 (1996).

Cardiac development

Transcription and the broken heart

Garry P. Nolan

Congenital heart defects occur in an astonishing 1% of live births^{1,2}, and fetal heart malformations are implicated in many pregnancies that end in still-birth or spontaneous abortion². Why, of all the body's complex organ systems, is the heart so susceptible to malformation during fetal development? Genetic factors are probably central to many congenital heart defects, but complicating factors such as environmental toxins or the diet and age of the parents are likely to affect the severity of these disorders. Despite compelling statistics on their incidence, and in the face of exciting genetic progress^{3,4} implicating several genes in the developmental programme of the normal heart, we still do not understand biochemically why heart defects are so prevalent. This is mainly due to a lack of the molecular targets that might indicate preventative therapeutic approaches or outline signalling systems for further investigation.

On pages 182 and 186 of this issue, de la Pompa *et al.*⁵ and Ranger *et al.*⁶ offer tantalizing hints as to how a single transcriptional regulator might link the genetic and environmental causes of one class of congenital heart disorders — birth defects involving valve and septum formation. Both groups specifically mutated a mouse gene that encodes a member of the NF-AT

(nuclear factor of activated T cells) family of transcription factors. This gene (*NF-ATc*) is known to act in the immune system⁷. But, unexpectedly, as a result of the introduced mutations, the pulmonary and aortic valves completely and specifically failed to form. These critical valves control the flow of blood from the ventricles of the heart into the arteries leading to the lungs and the main circulation, respectively (Fig. 1, overleaf).

The ventricular septal structure, which separates the left and right ventricles, was also rendered defective by the mutations. Less severely affected (although clearly defective in the study of de la Pompa *et al.*), were the tricuspid and mitral valves, which regulate the flow of blood from the right and left atria into their respective ventricles. The spectrum of malformations found in human hearts more closely resembles the defects seen in the pulmonary and aortic valves of the mouse *NF-ATc* mutants. Human clinical cases only rarely involve the complete lack of all four valves, consistent with the phenotypes observed in the mice. However, as a result of these combined defects observed in both studies^{5,6}, the mouse embryos died after 14–15 days' gestation.

One of the defects caused by disruption of the mouse *NF-ATc* gene is strikingly similar to a class of human congenital heart diseases. This class describes defects in

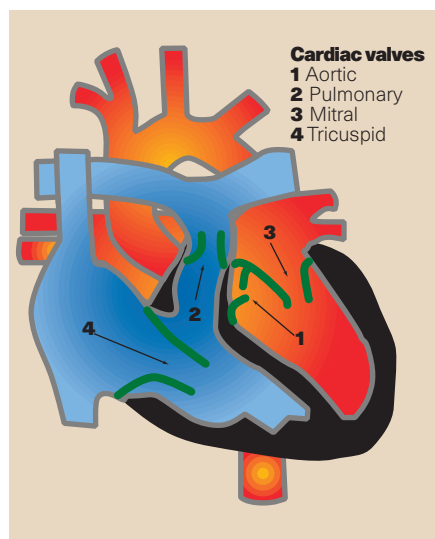


Figure 1 Cardiac valves that depend on the activity of fetal NF-ATc. The structure of the normal human heart is shown: the right atrium and ventricle are in blue; the left atrium and ventricle are in red. Blood from the main circulation enters the right atrium and passes through the tricuspid valve (4, green). The right ventricle pumps blood through the pulmonary valve (2) and into the lungs. Oxygenated blood returns from the lungs into the left atrium and passes through the mitral valve (3) into the left ventricle. From here, the ventricle pumps the blood through the aortic valve (1) and into the main circulation. de la Pompa *et al.*⁵ and Ranger *et al.*⁶ have found that the aortic and pulmonary valves are not formed in *NF-ATc*^{-/-} homozygous deletion mice, and de la Pompa *et al.* found that the mitral and tricuspid valves are also malformed.

formation of the atrioventricular valves and septal structures that, combined, separate the four chambers of the developing heart (Fig. 1). During development, the septal and valve structures originate from endocardial cells that migrate into the 'cardiac cushions' — regional swellings on the inner lining of the developing heart that eventually give rise to the septum and valves. Migration of these endocardial cells results in their transformation into mesenchymal cells, apparently in response to soluble protein signals. This transformation seems to be essential to the participation of endocardial cells in development of the valve leaflets or septal structures. A burst of *NF-ATc* activation is observed in the developing 'primordial' septal structures of the cardiac cushions in wild-type animals, correlating exactly with initiation of their formation. Thus, disruption of the *NF-ATc* gene interrupts a normal and important function for NF-ATc in the primordia.

Although de la Pompa *et al.*⁵ and Ranger *et al.*⁶ find that *NF-ATc* is expressed in endocardium and valve primordia, we would have predicted from the literature that the main

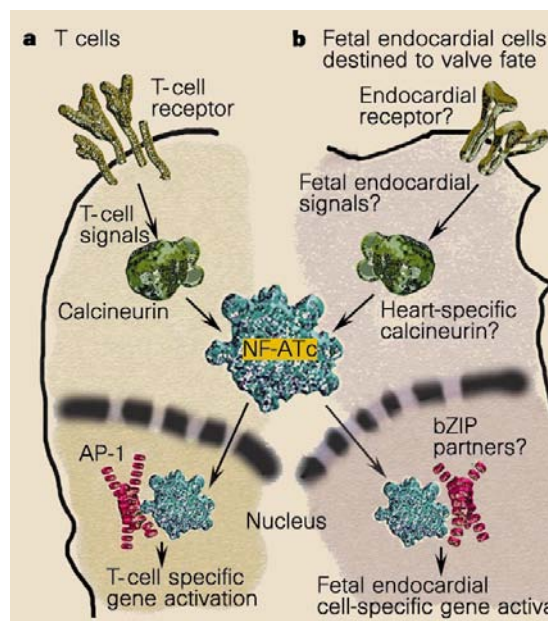


Figure 2 Common paths for activation of NF-ATc in T cells and fetal endocardial cells? a, T-cell activation involving NF-ATc. Engagement of the T-cell receptor (and other co-receptors) triggers the activation of the phosphatase calcineurin. Calcineurin dephosphorylates the amino terminus of NF-ATc, causing it to translocate into the nucleus, where it is thought to interact with newly synthesized proteins such as the bZIP protein AP-1 (Fos/Jun) and activate specific genes. b, A comparable activation system may be used in the developing fetal heart. Endocardial cells activate NF-ATc through a calcineurin-specific pathway. Upstream activators and downstream target loci are proposed by analogy to the well-characterized T-cell system.

effect of knocking out the gene would be immune dysfunction. This is because the NF-AT holoprotein complex was originally characterized by Crabtree and colleagues as a critical component of the transcriptional regulatory apparatus of interleukin-2 in antigen-stimulated T cells. NF-AT is composed of at least two components. In T cells, a cytoplasmic component pre-exists, containing a DNA-binding domain that is homologous to Rel. So far, four genes are known⁷ to encode proteins belonging to this subgroup — *NF-ATc*, *NF-ATp*, *NF-AT3* and *NF-AT4*. Although these genes are expressed in various tissues, their role is best understood in the immune system. After T-cell activation, the Rel component (for example, NF-ATc) translocates to the nucleus, where it can bind to DNA directly, or in a complex with members of the bZIP transcription-factor families^{7,8} (Fig. 2).

Amazingly, it seems that key aspects of these NF-ATc signalling processes are also used during early heart development. Moreover, genetic disruption of *NF-ATp* (which is biochemically and genetically similar in many ways to NF-ATc) leads to no obvious defect in heart development⁹. But it does disrupt the immune system, as might have been expected⁹. Because *NF-ATp* does not seem to be expressed in heart tissue, other transcription factors must control the temporal and tissue-specific expression of *NF-ATc*, distinct from that of *NF-ATp*.

How does NF-ATc respond to signals in endocardial and T cells? In part, it is thought to interpret changes in the levels of calcium in the cytoplasm. Calcineurin — a protein that responds to calcium signalling — is known to activate NF-ATc by dephosphorylating residues in its amino terminus. By blocking calcineurin function using pharmacological agents such as FK506 and cyclosporin A, activation of NF-ATc can be

blocked. Moreover, de la Pompa *et al.*⁵ and Ranger *et al.*⁶ showed that movement of NF-ATc from the cytoplasm to the nucleus in cells from these developing heart structures can be inhibited by FK506 and cyclosporin A, respectively.

It is tempting to speculate, therefore, that defects in cardiac valve formation are caused by problems in NF-ATc signalling. These may be genetically determined, or they could be influenced by environmental toxins that deregulate the heart's developmental programme. For example, if inappropriate environmental signals are present during the critical stage of valve and septal development in the human embryo, they could lead to cardiac birth defects. Moreover, other forms of valve defect, such as stenosis (narrowing of the valve orifice) and mitral valve prolapse (which affects many women in later life and is a primary risk for endocardial infection), might result from subtle deregulation of the signalling systems linked to NF-ATc function. So the upstream activators and target loci of NF-ATc in the developing heart should now be investigated, because the activity of these proteins — or their misregulation — could be crucial in eradicating this important class of birth disorders. □

Garry P. Nolan is in the Department of Molecular Pharmacology and the Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, California 94305, USA.
e-mail: gnolan@cmgm.stanford.edu

1. Ferencz, C. *et al.* *Am. J. Epidemiol.* **121**, 31–36 (1985).
2. Hoffman, J. I. E. *Pediatr. Cardiol.* **16**, 155–165 (1995).
3. Rossant, J. *Circulation Res.* **78**, 349–353 (1996).
4. Olsen, E. N. & Srivastava, D. *Science* **272**, 671–676 (1996).
5. de la Pompa, J. *et al.* *Nature* **392**, 182–186 (1998).
6. Ranger, A. M. *et al.* *Nature* **392**, 186–190 (1998).
7. Rao, A., Luo, C. & Hogan, P. G. *Annu. Rev. Immunol.* **15**, 707–747 (1997).
8. Nolan, G. P. *Cell* **77**, 795–798 (1994).
9. Xanthoudakis, S. *et al.* *Science* **272**, 892–895 (1996).

Biophysics

A cold break for photoreceptors

Lars-Oliver Essen and Dieter Oesterhelt

Bacteria, plants and animals all perceive light using photoreceptors, each of which carries a bound, organic molecule known as a chromophore. When hit by a photon, the chromophore undergoes an apparently simple reaction — a *cis*-to-*trans* isomerization around a double bond. Light-perception by the chromophore is ultra-fast, taking only a few hundred femtoseconds. But the ensuing structural changes of the chromophore and its protein environment are thermally driven and much slower. For example, it takes more than 10^8 times longer for a photoreceptor to reach its signalling state, generate a cellular signal and, finally, to complete the 'photocycle' by returning to its original state.

How do these photoreceptors translate femtosecond events into millisecond responses so efficiently? In the atomic-resolution study reported by Genick *et al.*¹ on page 206 of this issue, we get a first impression of a biological photoreceptor that has been trapped very shortly after excitation. The authors studied the photoactive yellow protein (PYP), which was discovered in the halophilic bacterium *Ectothiorhodospira halophila*². There, as well as in other phototrophic purple bacteria, PYP serves as a light sensor which, presumably, mediates the negatively phototactic response of these bacteria to blue light.

Originally, PYP was thought to be a small, soluble cousin of the intensively studied bacterial and animal rhodopsins. These proteins are embedded in lipid membranes, and they use the vitamin-A derivative retinal as the chromophore. But X-ray crystallographic work³ revealed that PYP has a completely different protein architecture from its rhodopsin-like ancestors³. It was also found that it carries the simplest photoisomerizable chromophore, *p*-hydroxycinnamic acid (*p*CA). This chromophore contains a single isomerizable double bond, flanked by a bulky phenolic ring on one side and a thioester linkage to PYP on the other side (Fig. 1).

Despite these structural differences, there are several common aspects that make PYP a fascinating model system for the retinal-containing rhodopsin family. First, the protein performs a large redshift for the absorption maximum of the *p*CA chromophore. Second, here are several, spectroscopically well-discriminated intermediates within the photocycle. Third, the triggering is extremely efficient — only one-third of the excited photoreceptors fail to initiate a full photocycle. Last, but not least, PYP is of considerable benefit to the structural biologist. It forms outstanding, high-quality crystals that are amenable to state-of-the-art synchrotron techniques, time-resolved Laue crystallography⁴ and studies on cryo-

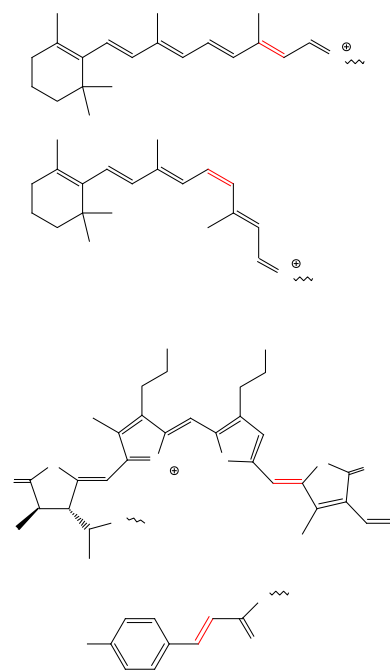


Figure 1 Chromophores from different photoreceptors. a, From archaeal rhodopsins. b, From animal rhodopsins. c, From plant and bacterial phytochromes. d, From bacterial photoactive yellow protein. The isomerizable double bond in each case is highlighted in red.

trapped intermediates at sub-ångstrom resolution, as now reported by Genick *et al.*¹.

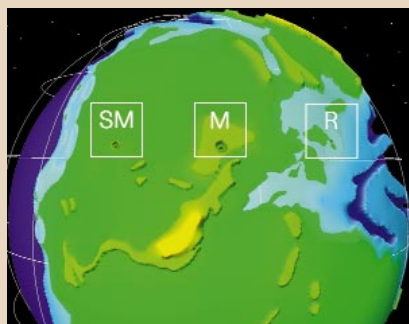
The time-resolved techniques⁵ use a laser flash to initiate the photoreaction in a crystal, followed by a second, 'white' X-ray pulse to record the structural information. Using a

Planetary science

Crater row

A crater chain is an alignment of three or more impact craters, with the same apparent age, thought to be caused by the fragmentation and sequential collision of large bodies with a planet. But finding such chains on the Earth's surface is complicated by the process of plate tectonics which, over geological time, will disperse the craters. Elsewhere in this issue (*Nature* 392, 171–173; 1998), John G. Spray and co-workers describe how they have identified just such a chain. It consists of at least five craters, spread over more than 4,000 km, which were produced about 214 million years ago. The illustration here shows the three main craters in the chain — St Martin, SM; Manicouagan, M; and Rochechouart, R — which line up at constant latitude when the Earth's tectonic plates are reconfigured to their arrangement at the time of the collisions.

In 1994, the break-up of comet Shoemaker–Levy 9 by the tidal forces of



Jupiter, and the subsequent collision of the fragments with the planet, dramatically showed how such crater chains might be formed. The fact that the craters observed by Spray *et al.* lie at a constant latitude indicates that the fragmented body that produced them had likewise been captured into an Earth orbit, with the distance between craters giving the angle through which the Earth rotated between impacts. But the large difference in mass between Jupiter and

the Earth (over 300 to 1), and the Earth's closer proximity to the Sun, mean that it seems unlikely that the Earth would have captured and, through tidal forces, fragmented a comet or asteroid in the same manner as Jupiter.

An alternative view of events is this — a comet or asteroid makes a grazing passage through the Earth's atmosphere; it is fragmented and the pieces are captured into an Earth orbit through the forces involved in 'aerobraking'; the fragments then collide with the planet at their next pass-by (see W. F. Bottke *et al. Icarus* 126, 470–474 (1997); S. G. Love *et al. Lunar Planet. Sci. Conf. XXVIII*, 837–838 (1997)). Such a process would still happen only infrequently. But it is estimated to have affected roughly 1 out of every 1,000 objects that have hit the Earth, and so could have been responsible for the crater chain observed by Spray and colleagues.

John VanDecar

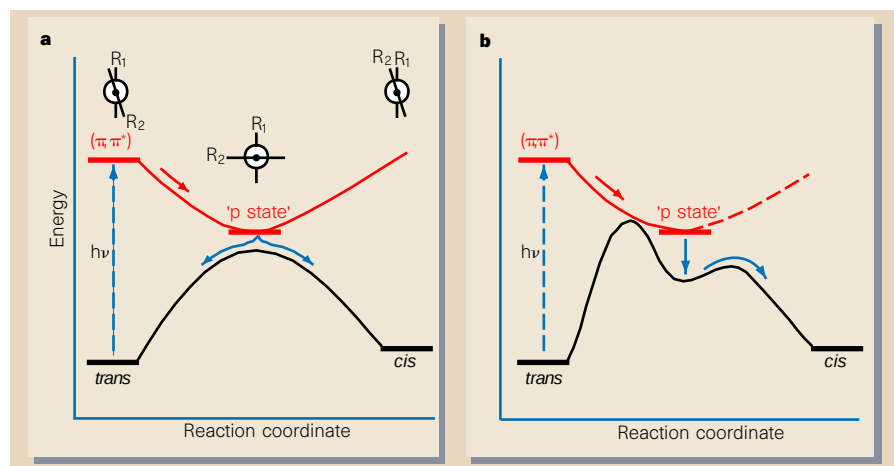


Figure 2 Energy diagrams for the photoisomerization of 1,2-disubstituted ethenes. **a**, In free solution. **b**, In the context of a photoreceptor. The energy profile of the excited (π, π^*) state (shown in red) is different than that for the thermal isomerization. Genick *et al.*¹ have observed that early-state photoreceptor intermediates of the photoactive yellow protein resemble this 'p-state' intermediate.

similar strategy, a 1.9-Å structure of the late, blueshifted I_2 PYP intermediate has been obtained⁴, several milliseconds after excitation. In this intermediate, the pCA chromophore flips its double bond to a *cis* configuration by swinging the phenolic ring like a tennis racket. This large movement causes a considerable reorganization of surface-exposed protein residues nearby, and might enable PYP to interact with other signalling molecules. The chromophore is not planar, and the authors supposed this to be the molecular spring for completion of the photocycle.

An absorbed photon supplies more than enough energy to generate such late intermediates. But how is the energy that is initially localized in a 'hot' chromophore converted to the gross structural changes of the protein environment? And how can a photoreceptor catalyse a single reaction pathway, avoiding the unwanted side reactions that are well known to photochemists? The answer might be found in the structures of very early intermediates. However, these are formed too rapidly to be studied by the time-resolved approach, which operates on a nanosecond timescale⁶.

Genick *et al.*¹ offer a simple but elegant solution to these problems. They excited crystals of PYP at liquid-nitrogen temperatures, and assumed that the degrees of freedom that were frozen out were those responsible for the slower thermal relaxations towards the I_2 intermediate. Their 0.85-Å structure of such an early PYP photocycle intermediate shows, for the first time, a chromophore that has started a *trans*-to-*cis* isomerization around its C7–C8 double bond, but has become trapped halfway between.

What does the orthogonal conformation of the double bond tell us about a mechanism for photoreceptors? Consider, as the simplest model, 1,2-disubstituted ethenes in free solu-

tion. When they are excited by light, they form a transient (π, π^*) state in which there is effectively no π bond. Subsequent rotation around the bond gives a staggered, lower-energy conformer in which the adjacent p-orbitals are perpendicular to one another (Fig. 2a). This relaxed 'p state' can decay, without emitting radiation, to a *cis* or *trans* isomer with almost equal probability, by crossing to the ground-state energy curve⁷.

Genick *et al.* observe that, in their 0.85-Å structure, the chromophore looks similar to such a p-state intermediate. One could speculate that photoreceptors maximize

their quantum efficiencies by selectively stabilizing p-state-like intermediates. This is analogous to the way in which enzymes are efficient catalysts — by selective stabilization of the transition state. For PYP, the non-uniform protein environment means that photoisomerization is unidirectional, because the energetic barriers for returning to the original configuration or progressing through the reaction are not equal (Fig. 2b).

To understand the small, structural changes in early photocycle intermediates, we certainly need more sub-ångstrom structures like that presented by Genick *et al.* Unfortunately, it will probably be a long time before we have such an eagle-eyed view of rhodopsins or phytochromes — the quality of these crystals is a long way behind that of PYP. But the new results on PYP mean that crystallographers, quantum chemists and spectroscopists have got good reasons to chat to one another. □

Lars-Oliver Essen and Dieter Oesterhelt are at the Max-Planck-Institut für Biochemie, Am Klopferspitz 18a, D-82152 Martinsried bei München, Germany.

e-mail: essen@biochem.mpg.de

- Genick, U. K., Soltis, S. M., Kuhn, P., Canestrelli, I. L. & Getzoff, E. D. *Nature* **392**, 206–209 (1998).
- Meyer, T. E. *Biochim. Biophys. Acta* **806**, 175–183 (1985).
- Borgstahl, G. E. O., Williams, D. R. & Getzoff, E. D. *Biochemistry* **34**, 6278–6287 (1995).
- Genick, U. K. *et al.* *Science* **275**, 1471–1475 (1997).
- Moffat, K. & Ren, Z. *Curr. Opin. Struct. Biol.* **7**, 689–696 (1997).
- Imamoto, Y., Kataoka, M. & Tokunaga, F. *Biochemistry* **35**, 14047–14053 (1996).
- Gilbert, A. & Baggott, J. in *Essentials of Molecular Photochemistry* 229–285 (Blackwell, Oxford, 1991).

Meteoritics

Diamonds in the dust

Primitive meteorites are time capsules, containing dust that may have remained virtually unchanged since the earliest days of the Solar System. The dust is often peppered with tiny diamonds, which probably condensed from a stellar outflow before the birth of the Sun. But uncertainty surrounds the types of star involved.

Anja Andersen and colleagues (*Astron. Astrophys.* **330**, 1080–1090; 1998) have investigated the matter by extracting diamonds from the 1969 Allende meteorite, shown here in greatly magnified clusters. They then analysed the diamonds' spectral and absorption characteristics at various wavelengths, and predict the spectrum of a plausible candidate star (in this case, a type of carbon-rich red giant). With a template against which to compare real stellar spectra, the origin of the diamonds could in principle be revealed.

Things are not so simple, however, for diamonds have a comparatively weak influence on stellar spectra. But other lines



of approach may open up because these microscopic gems can act as seeds, and may have become covered with coats of other dust grains. Andersen *et al.* suggest such coatings might be retained and their composition revealed by sensitive extraction techniques — offering further clues to the diamonds' history.

Karen Southwell



100 YEARS AGO

Not that the quest of birds' eggs is in any way to be discouraged – far from it. While the eggs themselves, from the intrinsic beauty of their shape and coloration, form a never-ending source of delight to the collector, it is only through the energy of the hunter after these spoils that a knowledge of the habits of the birds themselves can ever be attained. In the words of Prof. A. Newton, it is the field naturalist who alone crowns the labours of the anatomist and the museum worker. "What engineer," writes the Professor, "can be said to understand his business if he knows not the purpose to which the machines he makes are to be applied, and is unacquainted with their mode of working. ...". All honour then to those who risk their health, if not their lives, and spend their treasure, in the quest of the rarer birds' eggs!

From *Nature* 10 March 1898.

50 YEARS AGO

The real difficulty is that applied psychiatry involves two requisites: a technique to be acquired by a specific training, and a personal character which cannot as yet be acquired. An industrial physicist or chemist may be a drunkard, a wife-beater or a child murderer without it necessarily interfering with his scientific work; but an industrial psychiatrist can ill afford these idiosyncrasies. Whatever it means to-day, psychology used to mean the study of the human 'soul', and there used to be a profession or a vocation called a 'cure of souls'. Why the cure of souls has been taken over from the clergyman by the psychiatrist is a question which cannot be gone into here. Possibly there are too many clergy who have character without technique, but it is certain that technique by itself will be no adequate substitute for character. The admirable speech of the Earl of Feversham, chairman of the National Association [for Mental Health], left the conference in no doubt that the Association is fully alive to this necessity for wedding technique and character in all cure of souls, and along with Mr Edwards' opening address must have sent the delegates away realizing that they must act not only as administrators but also as human beings if the rising tide of mental sickness is to be held back.

From *Nature* 13 March 1948.

Superconductors

Opening the gap

Piers Coleman

The discovery that metals emit electrons when exposed to ultraviolet light led Einstein to the radical conclusion that light is made up of photons. Physicists' long affair with photoemission continues today. As the paper by Norman and colleagues on page 157 of this issue¹ shows, it is changing the way we think about the collective motions of electrons in a new class of metals — the high-temperature superconductors.

Electrons are fermions, so no two can share the same quantum state. This is what leads to the diversity of the periodic table. In metals, it means that electrons can't accumulate in the states of lowest energy and momentum, but instead spread out uniformly in momentum up to a particular maximum energy called the Fermi energy. In this way, electrons form a kind of droplet in momentum space (the Brillouin zone), with those on the droplet surface (the Fermi surface) being at the Fermi energy. The deformations and vibrations of the Fermi surface govern the collective properties of a metal.

High-temperature cuprate superconductors burst upon the world eleven years ago². These ceramics are a qualitatively new kind of metal: electrons are confined within conducting layers of copper oxide, so the materials conduct poorly in the third direction, and the Brillouin zone is two-dimensional. The many unusual properties of these metals seem to be a result of new types of correlated electron motion that develop long before superconductivity — motions that make their mark on the Fermi surface.

Superconductivity develops in metals when electrons at the Fermi surface bind together into pairs. These pairs are not governed by the fierce individualism of free electrons, so they can all drop into the lowest-energy quantum state, forming a pair condensate. Electrons in the condensate move without resistance, short-circuiting the resistance of those that have not condensed. It costs some energy to pull an electron out of the condensate — this is the superconducting energy gap. The gap stabilizes the condensate and thus is vital in the development of superconductivity. But few details of how the gap develops are known. That is where photoemission comes in.

Angle-resolved photoemission, or ARPES, is at present the only available way to probe the momentum of electrons in these unusual metals. When a high-energy photon scatters off the surface of a metal, it evicts an electron to form a 'hole' in the electron fluid. The intensity of scattered electrons provides a direct measure of the electron energy and

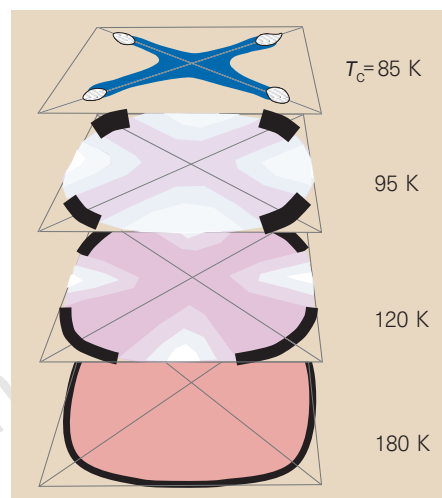


Figure 1 The Fermi surface in underdoped $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$. As the temperature is reduced, the surface evolves from a continuous rounded square at high temperatures, into four discontinuous arcs, which shrink into four nodes below the superconducting transition temperature of 85 K.

momentum distribution.

Two groups, led by Z.-X. Shen at Stanford University, and by J. C. Campuzano at the University of Illinois and Argonne National Laboratory, have pioneered ARPES measurements on the cuprate superconductors. These materials come in three basic varieties, in which the amount of oxygen or other dopant is progressively increased: underdoped; optimally doped (where the superconducting transition temperature is at a maximum); and overdoped.

ARPES was first used on the optimally doped materials, where it reveals a Fermi surface with a rounded-square cross-section^{3,4}. It also shows how the energy gap develops. Unlike that in ordinary metals, the gap depends strongly on momentum and vanishes at the corners of the Fermi surface⁵. This is some of the most direct evidence we have for the idea that high-temperature superconductivity is 'd-wave' superconductivity. In such a superconductor, the wavefunction of each pair changes sign four times around the Fermi surface. This sign alternation keeps the electrons in each pair far apart, and it is thought to be this correlated motion that lowers the energy of the condensate relative to the normal metal. It also causes the energy gap to vanish at the corners of the Fermi surface, where the pair wavefunction passes through zero.

But the underdoped cuprates deviate even further from conventional metallic behaviour. Some, such as the bilayer com-

pound $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$ (or Bi2212), appear to have a 'pseudogap' in their excitation spectrum — a depletion of electrons on the Fermi surface. The pseudogap develops long before superconductivity, indicating that some new kind of correlated motion is developing within the electron fluid. But is the pseudogap caused by superconducting correlations that begin to develop in the normal state^{6–8}? Or is it an independent phenomenon, perhaps a 'spin gap' caused by correlations between the orientation of spins in adjacent copper-oxide layers^{9–11}?

The first ARPES studies of underdoped Bi2212 showed that the pseudogap appears principally on the flattened faces of the Fermi surface, where its effect on the spectrum is almost identical to that of fully developed superconductivity^{12–14}. These early results showed that the Fermi surface remained intact at the corners.

The latest results confirm this, and show in a lot more detail what is going on as the temperature falls. Norman *et al.*¹ see the pseudogap develop on the faces of the Fermi surface, breaking it up but leaving the shape of the remaining arcs unchanged (Fig. 1). As the temperature is lowered further, each arc is progressively eaten away, finally vanishing at the superconducting transition temperature.

This is a controversial exercise, for many aspects of the temperature dependence of ARPES spectra are poorly understood even in conventional materials. Nevertheless, the new results add credence to the view that the pseudogap is a prelude to superconductivity. In the simplest view, electrons on the Fermi surface fleetingly come together as pairs long before superconductivity is fully developed. The results do not rule out another interpretation, but they demand that the correlations that produce the pseudogap should be the very same ones that form the superconductor. Electron-tunnelling measurements¹⁵ on Bi2212 point in the same direction, showing that the pseudogap in the normal state evolves smoothly into the superconducting gap.

These results can be quite simply understood¹⁶ in terms of virtual d-wave pair formation: as the temperature is lowered, the pairs decay more slowly. According to the energy uncertainty principle, the lifetime Δt of a pair is inversely proportional to the uncertainty in the pair binding energy, $\Delta\epsilon = \hbar/\Delta t$. When this uncertainty in the binding energy becomes comparable to the size of the d-wave superconductivity gap at a particular point on the Fermi surface, a pseudogap opens up. As the d-wave gap is large on the faces but vanishes on the corners, this naturally leads to the development of arcs at the corners, which shrink and eventually disappear.

More experiments will be needed to confirm or refute this interpretation. One vital question is why the pseudogap has not been

seen in all cuprate superconductors: is this a matter of resolution, or a fundamental difference between materials? The electron-tunnelling experiments¹⁵ clearly show that the pseudogap can be present, albeit reduced, in overdoped compounds, and a pseudogap has also been seen in a single-layer compound¹⁷. So perhaps the pseudogap is a feature of all high-temperature superconductors.

The true nature of the curious metal that gives birth to high-temperature superconductivity is still obscure. But almost everyone agrees that electron physics is undergoing a revolution, one in which experiments like that of Norman *et al.* will be the final arbiters. □

Piers Coleman is in the Department of Physics and Astronomy, Rutgers University, 136 Frelinghuysen

Road, Piscataway, New Jersey 08854-8019, USA.

e-mail: coleman@physics.rutgers.edu

1. Norman, M. R. *et al.* *Nature* **392**, 157–160 (1998).
2. Batlogg, B. *Phys. Today* **6**, no. 44, 44–50 (1991).
3. Olson, C. G. *et al.* *Science* **245**, 731–733 (1989).
4. Campuzano, J. C. *et al.* *Phys. Rev. Lett.* **64**, 2308–2312 (1990).
5. Shen, Z.-X. *et al.* *Phys. Rev. Lett.* **70**, 1553–1556 (1993).
6. Doniach, S. & Inui, M. *Phys. Rev. B* **41**, 6668–6678 (1990).
7. Randeria, M., Trivedi, N., Moreo, A. & Scalettar, R. *Phys. Rev. Lett.* **69**, 2001–2004 (1992).
8. Zachar, O., Kivelson, S. A. & Emery, V. J. *Phys. Rev. Lett.* **77**, 1342–1345 (1997).
9. Millis, A. J. *et al.* *Phys. Rev. B* **53**, 415–424 (1996).
10. Anderson, P. W. *J. Phys. Condens. Matter* **8**, 10083–10088 (1996).
11. Wen, X. G. & Lee, P. A. *Phys. Rev. Lett.* **76**, 503–506 (1996).
12. Marshall, D. S. *et al.* *Phys. Rev. Lett.* **76**, 4841–4845 (1996).
13. Loeser, A. G. *et al.* *Science* **273**, 325–329 (1996).
14. Ding, H. *et al.* *Nature* **382**, 51–53 (1996).
15. Renner, Ch. *et al.* *Phys. Rev. Lett.* **80**, 149–152 (1998).
16. Norman, M. R., Randeria, M., Ding, H. & Campuzano, J. C. <http://xxx.lanl.gov/archive/cond-mat/9711232>
17. Harris, J. M. *et al.* *Phys. Rev. Lett.* **79**, 143–146 (1997).

Membrane sorting

Endosome marker is fat not fiction

Sandra L. Schmid and Pieter R. Cullis

Each organelle in a eukaryotic cell performs a specific set of functions that are reflected in the unique protein composition of that organelle, and such organelle-specific or 'marker' enzymes have been invaluable in monitoring the purification of organelles by subcellular fractionation. Appropriately labelled antibodies directed against organelle-specific marker proteins have also been used to identify organelles and study their structure by immunofluorescence or immuno-electron microscopy.

On page 193 of this issue Kobayashi and colleagues¹ report how, in characterizing a new monoclonal antibody that specifically identifies late endosomes, they unexpectedly found that the antigen recognized is a lipid — lysobisphosphatidic acid (LBPA). The presence of the antibody perturbs both the structure and function of the late endosomes, suggesting a critical role for LBPA. Moreover, the authors found that this same lipid antigen is recognized by autoimmune

sera from patients with antiphospholipid syndrome, providing new insights into the pathological basis of this disease.

Lysosomes are the digestive organelles of the cell — they have highly acidic contents, rich in hydrolytic enzymes. Macromolecules such as nutrients, growth factors and foreign antigens are captured by receptors on the cell surface, for uptake and delivery to lysosomes via the endocytic pathway². Other intracellular receptors, particularly the mannose-6-phosphate receptor (MPR), capture and divert hydrolytic enzymes from biosynthetic pathways to the lysosomes. For efficiency, both the endocytic and biosynthetic receptors must be retrieved before they encounter the hostile environment of the lysosome.

Endosomes are common intermediates along both the biosynthetic and endocytic pathways to the lysosome (Fig. 1, overleaf). They are a structurally diverse population of vacuoles and tubules, and they serve as sorting stations for the retrieval of recycling receptors.

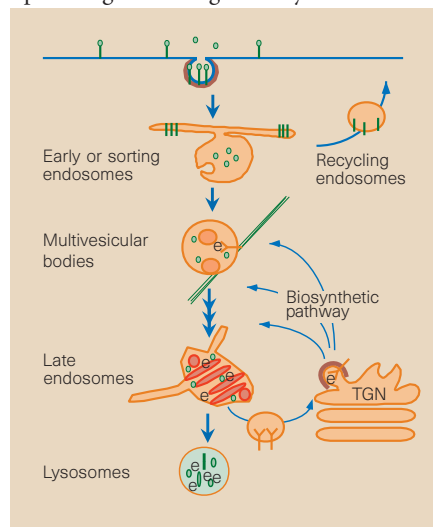


Figure 1 Endosomes — sorting stations along the endocytic pathway. Cell-surface receptors capture extracellular nutrients and other macromolecules (green spots), whereas receptors in the trans-Golgi network (TGN) capture inactive precursors of hydrolytic enzymes (e), for eventual delivery to lysosomes via the endosomal intermediates. Before the contents of the endosomes are delivered to lysosomes — where the hydrolases are activated and degradation occurs — the receptors must be recycled. Endosomes carry out this sorting and recycling function as they mature and accumulate intraluminal membrane. Kobayashi *et al.*¹ have found that a unique lipid and autoantigen, lysobisphosphatidic acid, is specifically localized to the intraluminal membrane of late endosomes and lysosomes. Its function there remains unknown.

Most cell-surface receptors are recycled from 'early' or 'sorting' endosomes (that is, those first encountered by endocytic tracers) where the pH is mildly acidic, favouring dissociation of the cargo from receptors. The released luminal contents are collected in large vacuolar portions of sorting endosomes, and the receptors are laterally segregated into long tubular extensions. The vacuoles then dissociate and move along microtubules towards the centre of the cell, where they start to mature. Maturation involves the removal of residual, recycling surface receptors, delivery of lysosomal hydrolases, and involution of the surrounding membrane to form multivesicular bodies³.

At the end of this process, the so-called 'late' endosomes contain large amounts of intraluminal membrane, and they are enriched in lysosomal hydrolases and lysosomal membrane proteins. Having delivered their cargo of lysosomal hydrolases, MPRs are removed and recycled to the trans-Golgi network before the late endosomes fuse with the lysosomes. The mechanisms underlying maturation of the late endosomes and the sorting and retrieval of recycling receptors are poorly understood. Perplexingly, the recycling MPRs are concentrated on the intraluminal membrane, whereas lysosomal membrane proteins are found on the limiting membrane of the late endosome⁴.

Cell biologists are only beginning to understand why cellular membranes contain so many types of phospholipid, and what particular species of lipid do in signal transduction, vesicle formation and fusion, and membrane protein-sorting⁵⁻⁷. The finding that LBPA is a 'marker' lipid of late endosomes (and lysosomes⁸) raises the question, what late-endosome-specific function might this unique lipid fulfil? Its structure may provide some clues. The small, negatively charged headgroup and polyunsaturated acyl-chain composition⁹ would be expected to result in physical properties similar to those of cardiolipin, which is a marker lipid in the highly convoluted inner mitochondrial membrane. Cardiolipin tends to adopt the non-bilayer lipid structures that might facilitate the mixing of lipid bilayers required for membrane fusion.

There are several possibilities for the function of LBPA. It may play a structural role in the maturation of late endosomes — its tendency not to form a bilayer could help in developing the complex intraluminal membrane system. Alternatively, its unique structure means that LBPA is resistant to phospholipases, so it may stabilize the late-endosome/lysosomal membranes against degradation. In this case, however, we might expect to find LBPA on external, rather than internal, membranes.

LBPA may also be involved in protein sorting by the late endosomes. Indeed, other species of lipid with long-chain fatty acids,

such as sphingomyelins, can form microdomains in the plane of the membrane that is implicated in protein sorting in polarized cells⁶. The exclusive localization of LBPA, within the internal membranes of late endosomes, suggests that these membranes represent a specialized functional domain. Furthermore, Kobayashi *et al.*¹ have shown that uptake of anti-LBPA antibodies into late endosomes perturbs the normal recycling of the MPR back to the trans-Golgi network. And the closely related lipid semi-LBPA has been shown¹⁰ to be enriched in tubular vesicular elements of the trans-Golgi network, where protein sorting is known to occur.

Late endosomes/lysosomes are degradative organelles, so LBPA may be involved in lipid catabolism. For example, fatty acyl-transferases¹¹ for LBPA exist in lysosomal fractions from rat liver, so LBPA could be involved in the sequestration and eventual transport of fatty-acid catabolites generated by lysosomal phospholipases. The effect of anti-LBPA antibodies on late endosomes — accumulation of membranes — is suggestive of other lysosomal storage diseases resulting from the aberrant accumulation of digestive products. Whether perturbation of any of these functions of LBPA accounts for the symptoms associated with antiphospholipid syndrome remains to be seen.

The work of Kobayashi *et al.* adds to the growing appreciation of the part played by lipids in membrane transport and sorting. Because the structure of lipids can be rapidly altered by the modification or removal of headgroups, or by acyl-exchange reactions, their associated functional properties can be tightly regulated. And given that cells contain over 200 unique species of lipid, it is not unlikely that others will turn out to be 'marker' lipids, involved in the structure and function of specific organelles. □

Sandra L. Schmid is in the Department of Cell Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA.

e-mail: slschmid@scripps.edu

Pieter R. Cullis is in the Department of Biochemistry, University of British Columbia, 2146 Health Sciences Mall, Vancouver, British Columbia, Canada V6T 1Z3.

e-mail pieterc@unixg.ubc.ca

1. Kobayashi, T. *et al.* *Nature* **392**, 193–197 (1998).
2. Mellman, I. *Annu. Rev. Cell Dev. Biol.* **12**, 575–625 (1996).
3. Futter, C. E., Pearce, A., Hewlett, L. J. & Hopkins, C. R. *J. Cell Biol.* **132**, 1011–1023 (1996).
4. Griffiths, G., Hoflack, B., Simons, K., Mellman, I. & Kornfeld, S. *Cell* **52**, 329–341 (1988).
5. DeCamilli, P., Emr, S. D., McPherson, P. S. & Novick, P. *Science* **271**, 1533–1539 (1996).
6. Simons, K. & Ikonen, E. *Nature* **387**, 569–572 (1997).
7. Chernomordik, L., Kozlov, M. M. & Zimmerberg, J. *J. Membr. Biol.* **146**, 1–14 (1995).
8. Wherrett, J. R. & Huterer, S. *J. Biol. Chem.* **247**, 4144–4120 (1992).
9. Stremmel, W. & Debuch, H. *Hoppe-Seyler's Z. Physiol. Chem.* **357**, 803–810 (1976).
10. Cluett, E. B., Kuusimäen, E. & Machamer, C. E. *Mol. Biol. Cell* **8**, 2233–2240 (1997).
11. Huterer, S. J. & Wherrett, J. R. *Biochem. Cell Biol.* **68**, 366–372 (1990).

Daedalus

Breathable water

When animals came out of the sea, they had to learn to breathe air. In fact, says Daedalus, they never did. They took their own water with them. Even today, we still breathe water: we absorb the oxygen which dissolves in the moist lining of our lungs.

So why can't we still breathe water? Sadly, bulk water contains too little oxygen for our modern needs — only about 0.7% by volume, compared to 21% in air. But Daedalus has a way out. Imagine, he says, a dense foam of tiny air-bubbles in water. If all the bubbles had the same diameter, and were packed closely together, they could not easily rise to the surface. The result would be a curiously viscous stable foam, analogous to those 'rigid' close-packed emulsions of oil in water, whose droplets can hardly move past one another.

To make this foam breathable, Daedalus will dissolve suitable salts in it, bringing it into osmotic balance with lung tissue, and will add the natural polysaccharides that give saliva and sputum their viscosity, and a detergent like the one that helps the lungs to expand. He will aerate the solution through a battery of uniform nozzles, compress it briefly to collapse bubbles smaller than the standard size, and drain it to pack the remainder tight. It will then contain 74% of air by volume, giving it a density of 0.26 g ml⁻¹.

Daedalus's 'Liquid Air' will be a novel environment. You will sink into it, but will still feel somewhat buoyed up. It will be easier to move through than water, more opaque and sound-deadening than the densest fog, and rather an effort to breathe. Extra oxygen may be needed to make it feel comfortable and safe. But then the novelty of the experience, the sense of entering a silent, private fluid world, should make the Liquid Air immersion bath a popular relaxation.

Other more serious uses should also develop. With its high water content, Liquid Air will be utterly fireproof. Pumped in volume from fire-tenders, it will blanket rescuers as they enter burning buildings; injected into threatened aircraft, it will extinguish fires and blind and disorientate hijackers. Even better, bullets and explosion-fragments would be rapidly halted by Liquid Air. Pumped into and around suspicious vehicles and packages, it would damp an explosion wonderfully. The blast would expend its energy in 'inverting' the air–water emulsion to a dense spray of liquid droplets.

David Jones

Ernst's ego

Max Ernst delved into his childhood experiences to find images for his art that would both explore Freudian psychology and mock those who put unquestioning faith in scientific rationality.

Martin Kemp

The charting of the darkly mysterious labyrinths of the unconscious mind in early twentieth-century psychology, above all by Freud, seemed at once to promise a new science of human nature and to threaten the intellectual structures of other institutionalized sciences as the true models of knowledge and as the secure foundations of modern life. The double-edged nature of the psychological sword was explored more profoundly by Max Ernst than by any other artist of his era.

Educated at the University of Bonn in Germany in 1909–12, Ernst was introduced to the experimental psychology of Wilhelm Wundt and his followers. Throughout his life, the artist's grasp of contemporary psychology extended far beyond the bowdlerized Freudian formulas favoured by many of his Surrealist colleagues.

The prime vehicle Ernst exploited to evoke the *mêlages* of memory images and the symbolic fantasies of dream states was the collage. He cannibalized, juxtaposed and overlaid images garnered voraciously from a wide range of general and specialist publications to paint the enigmatic landscapes of his inner mind. The popular science periodical, *La Nature*, was a favourite source.

The resulting juxtapositions may appear irrational yet Ernst's method was avowedly "experimental". "Collage is a hypersensitive and rigorously exact instrument, a seismograph capable of registering the exact potentialities of human welfare in every epoch," he wrote.

In series of collages from the early 1930s, pictures within pictures are "presented" by Loplop, the semi-mythical bird of his childhood, a pet cockatoo that had died at the moment of his sister's birth. This "*oiseau supérieur*" appears as a sinister caricature above the central image, and stands as his "alter ego", into which he displaces himself as a "third" person. Ernst was undoubtedly aware of Freud's *A Childhood Memory of Leonardo da Vinci*, which centred on Leonardo's infant memory of his mouth being touched by a bird's tail. Freud illustrated the 1919 edition with Oskar Pfister's fantastical diagram of a vulture concealed in the skirts of St Anne in Leonardo's painting in the Louvre.

Ernst's *Facility, also, Loplop Presents* combines images cut from an anatomical atlas with marbled paper and a blotter with scribbled sums, set against a drawn background.



Ernst's *Facility, also, Loplop Presents*, collage and pencil, 1931, private collection, London.

The space of the inner picture is systematically laid out in geometrical perspective – a reference to facile academic naturalism – while the ground occupied by Loplop's disembodied head offers no such coherence. The pedestrian 'science of art' as a mirror of scientific rationality is mocked by the leering bird.

Specialized scientific imagery often seemed in Ernst's eyes to be oddly mocking of its own claims. Ernst found that "anatomical or physical demonstrations... united such mutually distant physical elements that

the very absurdity of the array called forth in us a hallucinating succession of contradictory images, superimposed on one another with the persistence and rapidity of remembered lovemaking".

From Ernst's perspective, modern sciences and technologies, with their array of weird visual images, were manifestations of the strange obsessions of our psyche, and were seen as subject to the "uncertainty principle" that bedevils all acts of human observation and representation. □

Martin Kemp is in the Department of the History of Art, University of Oxford, 35 Beaumont Street, Oxford OX1 2PG, UK.

GroE is vital for cell-wall synthesis

Chaperone proteins help other proteins to fold. GroEL, the *Escherichia coli* form of the ubiquitous Cpn60 chaperonins, has a multimeric barrel-shaped structure with a central cavity, within which almost any protein can fold *in vitro*¹. But what does GroE (GroEL plus its co-chaperone GroES) fold in the cell? Why is it needed for cell survival? We report here the first definite identification of an essential, GroE-dependent *E. coli* protein, dihydropicolinate synthase (DapA), without which cell-wall synthesis fails.

Although GroE is abundant, calculation³ suggests there is enough in a cell to interact with only about 10% of newly synthesized proteins. While several hundred different species of newly synthesized protein transiently associate with GroE, they comprise fewer than 20% of possible types⁴, only a minority of which are likely to be essential.

We find that cells deprived of GroE lyse because they fail to make cell walls. Lysis is due to a lack of the cell-wall precursor diaminopimelic acid (DAP), which fails to be made because the level of DapA — the first enzyme in the DAP-synthesis pathway — is greatly reduced in GroE-deficient cells.

When GroE levels were reduced⁵ or a mutant form of the protein inactivated by temperature shift⁶, changes in the levels of many proteins were noted, but no particular protein was identified as both GroE-dependent and essential for cell survival. By replacing the native *groE* chromosomal promoter region with the *araC* gene and *araBAD* promoter (Fig. 1a), we created a strain dependent on added arabinose for continued viability and GroE production, and were able to use it to determine the cause of death as GroE is depleted.

Following arabinose removal, GroE levels halve with each generation (Fig. 1b). Growth remains exponential for about 2 hours (Fig. 2a; GroE level will be reduced to 2–3% of normal by this time) and then continues at a decreasing rate for another 1.5 hours; the cells then abruptly lyse. That cells can continue to grow exponentially with greatly reduced levels of GroE suggests either that any essential GroE-dependent proteins are present in considerable excess, or that they can effectively co-opt any GroE available³.

During GroE depletion in osmotically protected medium, the cells are converted to spherical protoplasts, indicating a defect in synthesis or maintenance of the peptidoglycan cell wall. We followed peptidoglycan synthesis by measuring the rate of incorporation of [³H]DAP, a specific precursor, into peptidoglycan⁷. Uptake was greatly enhanced in GroE-depleted cells, suggest-

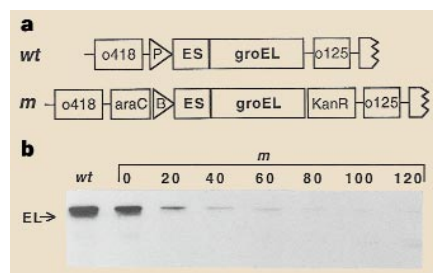


Figure 1 Depletion of GroE in *E. coli*. **a**, Genetic organization of the *groESL* region. wt, the native *groE* operon and surrounding open-reading frames. P represents the native *groE* promoters (σ_{32} and σ_{70}) and ES the *groES* gene. m, The *groE* region in the 'depletion mutant' strain. The *araC* gene and the arabinose-dependent *P*_{BAD} promoter (B) replace the native promoter; a kanamycin-resistance cassette is downstream of *groE*. **b**, GroEL depletion after arabinose removal. Mutant cells grown in L-broth at 37 °C with 0.1% arabinose were centrifuged, washed, diluted into broth with 0.2% glucose (time zero) and growth followed at 37 °C. Extracts of aliquots taken between 0 and 120 minutes were electrophoresed, electroblotted onto nitrocellulose, and GroEL (EL) was detected by chemiluminescence using polyclonal antiserum. An extract of the parental wild-type strain, MG1655, is included (wt). The mutant strain grown with 0.1% arabinose produces ~80% as much GroEL as does MG1655.

ing that these cells might be DAP-starved.

In confirmation of this, we found that addition of DAP to the growth medium (Fig. 2a) prevents cell lysis and permits growth to continue for up to 6 hours. Furthermore, although addition of DAP to solid medium does not support colony formation — confirming that GroE is essential for processes other than cell wall synthesis — when DAP is present, colony formation requires less arabinose than is normally the case. This suggests that at least a part of the cellular complement of GroE is dedicated to ensuring that DAP is available, and that GroE is likely to have a role in maintaining the levels of an enzyme or enzymes involved in DAP synthesis.

DAP is synthesized from aspartic semi-aldehyde in six steps⁸. Deficiency of any of the enzymes DapA–E causes cell lysis⁹. We reasoned that if a critical enzyme were to be synthesized in excess before GroE depletion, lysis would be delayed. We cloned *dapA*, *B*, *D* and *E* separately into pBR325 to achieve overproduction of the encoded enzymes. The presence of the *dapA*, but not of the other *dap* plasmids, delays lysis by several hours (Fig. 2a), suggesting that levels of DapA may be influenced by GroE. Immediately before lysis, the level of DapA in the depleted cells is reduced to 16% of normal (Fig. 2b). The remaining protein is soluble

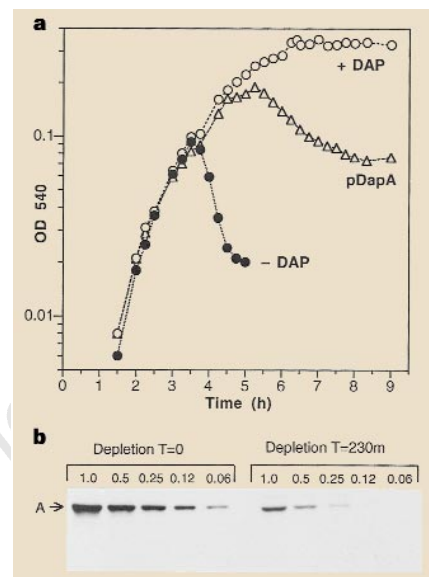


Figure 2 DapA loss in GroE-depleted cells. **a**, Cells depleted of GroE lyse (–DAP); addition of diaminopimelic acid (+DAP) prevents lysis. Lysis is delayed in cells in which DapA is overproduced (pDapA). The mutant was grown as for Fig. 1b, with or without DAP (50 μ g ml^{–1}). *dapA* was amplified by polymerase chain reaction and cloned into pBR325 (pDapA). pDapA was then transformed into the mutant. **b**, Levels of DapA decrease with GroE depletion. Mutant cells were grown and extracts prepared at the indicated times, as for Fig. 1b. Samples containing equal amounts of protein were serially diluted as indicated. Western blots were prepared using anti-DapA antiserum (gifts from G. Gallili and C. Falco).

and stable. In contrast, the levels of several other essential proteins examined do not decrease during GroE depletion.

Thus we see that levels of an essential protein, DapA, are dependent on GroE. We suggest that decreased stability of nascent DapA, as a consequence of failure of GroE-assisted folding, is the most likely reason for DapA deficiency.

Neil McLennan, Millicent Masters

*Institute of Cell and Molecular Biology,
Edinburgh University, Kings Buildings,
Mayfield Road, Edinburgh EH9 3JR, UK
e-mail: M.Masters@ed.ac.uk*

- Martin, J. & Hartl, F. U. *Curr. Opin. Struct. Biol.* 7, 41–52 (1997).
- Fayet, O., Ziegelhoffer, T. & Georgopoulos, C. *J. Bacteriol.* 171, 1379–1385 (1989).
- Lorimer, G. H. *FASEB J.* 10, 5–9 (1996).
- Ewalt, K. L. et al. *Cell* 90, 491–500 (1997).
- Kanemori, M., Mori, H. & Yura, T. *J. Bacteriol.* 176, 4235–4242 (1994).
- Horwich, A. L. et al. *Cell* 74, 909–917 (1993).
- Höltje, J. V. et al. *J. Bacteriol.* 124, 1067–1076 (1975).
- Cohen, G. N. in *Amino Acids: Biosynthesis and Genetic Regulation* (eds Herrmann, K. M. & Somerville, R. L.) 147–171 (Addison-Wesley, Reading, MA, 1983).
- Patte, J.-C. in *Escherichia coli and Salmonella: Cellular and Molecular Biology*, 2nd edn (ed. Neidhardt, F. C.) 528–541 (ASM, Washington DC, 1996).

Fluid 'rope trick' investigated

Buckling instabilities can arise from competition between axial compression and bending in slender objects. These are not restricted to solids, but also occur with fluids with free surfaces^{1–4}, in geophysics⁵ and in materials processing⁶. Here we consider a classic demonstration of fluid buckling⁷.

When honey is poured from a sufficient height, it approaches one's toast as a thin filament which whirls steadily around the vertical forming a regular helical coil (illustrated with silicone oil in Fig. 1), a behaviour reminiscent of the coiling of a falling flexible rope⁸. We derive a scaling law that predicts the coiling frequency in terms of the filament radius and the flow rate.

The physical parameters governing the phenomenon include the fluid density ρ , viscosity μ (kinematic viscosity $\nu = \mu/\rho$), the flow rate Q , gravity g , a characteristic filament radius r , and the height, h , over which the filament falls. As h is gradually increased, the axial stagnation flow becomes unstable to bending disturbances and the filament is steadily laid out in a circular coil of radius R at a frequency Ω .

The onset of the instability^{9–11} is determined by the relative magnitude of the gravitational timescale $(h/g)^{1/2}$ and the viscous timescale $r^2\rho/\mu$, characterized by a slenderness ratio $\epsilon = r/h$, or a Reynolds number $Re = gr^3/\nu^2$. We note that a similar parameter $\rho gh^3/B$, where B is the bending stiffness, occurs in the elastic analogue⁸. Only below a critical value of ϵ or Re is the axisymmetric stagnation flow unstable. Far from onset, when ϵ becomes sufficiently

small, the flow is still mainly an axial stretching flow. However, in a small neighbourhood of the flat surface there is a highly nonlinear coiling region, which persists over a range of falling heights.

In this region, the filament radius is constant (Fig. 1), and the rotatory inertial forces due to the whirling dominate gravity and are balanced by the viscous forces due to the differential velocities between the inner and outer segments of a curved filament. The characteristic radius of curvature of the filament scales with the coil radius R , as for a coiling rope⁸. Then the differential velocity scales as Ur/R , where U is the axial velocity. The Newtonian viscous force per unit volume scales as $\mu Ur/R^3$. The force per unit volume due to centripetal and Coriolis accelerations scales as $\rho\Omega^2 R$. Balancing the two forces yields

$$\mu Ur/R^3 \sim \rho\Omega^2 R \quad (1).$$

A filament of almost constant diameter is laid out in a steady circular coil near the flat surface $U \sim \Omega R$, whereas continuity of the axial stretching flow yields $Ur^2 \sim Q$. Substituting these relations into equation (1) yields a scaling law for the coiling frequency

$$\Omega \sim Q^{3/2} r^{-7/2} \nu^{-1/2} \quad (2).$$

To test these predictions, we studied the coiling of silicone oil, flowing at a constant rate out of circular holes in a steel plate at the base of a large reservoir of fluid. We established the dependence between the coiling frequency and filament radius by varying the height of the fall. The coiling frequency was determined by measuring the intensity fluctuations of a refracted laser aimed at the top of the coiling region. A horizontal microscope with a charge-coupled device camera was used to measure the filament radius with an accuracy of 1%. All measurements were made in a parameter regime far from the onset of coiling and also far from the regime when the coiled column itself is unstable and collapses periodically under its own weight. The data are best fitted by a power law $\Omega/Q^{1/5} \sim r^{-3.58 \pm 0.16}$

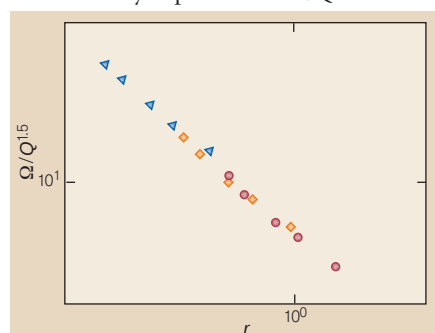


Figure 2 Log-log plot of the normalized coiling frequency ($\Omega/Q^{1/5}$) against filament radius (r). Triangle: hole radius, 3.2 mm, $Q=0.56 \text{ g s}^{-1}$; diamond: hole radius, 4.0 mm, $Q=0.96 \text{ g s}^{-1}$; circle: hole radius, 4.8 mm, $Q=1.28 \text{ g s}^{-1}$. The data are fitted by the relation $\Omega/Q^{1/5} \sim r^{-3.58 \pm 0.16}$ for each flow rate, and confirm equation (2).

for each flow rate (Fig. 2), and agree well with the theoretically predicted scaling law².

Several effects such as surface tension, the relaxation of Poiseuille flow to plug flow in the neighbourhood of the orifice, air drag, non-Newtonian effects and the effect of gravity in the coiling region, have been neglected here, as they are relatively unimportant. They can be accounted for in a quantitative long-wavelength theory similar to that used to describe coiling of falling ropes⁸.

L. Mahadevan

Division of Mechanical and Materials, Mechanical Engineering, 1-310, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
e-mail: l_m@mit.edu

William S. Ryu, Aravinthan D.T. Samuel

Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138, USA and Rowland Institute of Science, Cambridge, MA 02142, USA

1. Barnes, G. & Woodcock, R. *Am. J. Phys.* **26**, 205–209 (1958).
2. Taylor, G. I. *Proc. 12th Intl Congr. Appl. Mech.* 382–395 (1969).
3. Benjamin, T. B. & Mullins, T. J. *Fluid Mech.* **195**, 523–540 (1988).
4. Matovich, M. A. & Pearson, J. R. A. *Ind. Engrg. Chem. Fund.* **8**, 605–609 (1969).
5. Johnson, A. M. & Fletcher, R. C. *Folding of Viscous Layers* (Columbia University Press, New York, 1994).
6. Rawson, H. *Phys. Tech.* **2**, 91–114 (1974).
7. Taylor, G. I. *Low Reynolds Number Flows* (Encyclopaedia Britannica films, 1967).
8. Mahadevan, L. & Keller, J. B. *Proc. R. Soc. Lond. A* **452**, 1679–1694 (1996).
9. Cruickshank, J. O. & Munson, B. R. *J. Fluid Mech.* **113**, 221–239 (1981).
10. Griffiths, R. W. & Turner, J. S. *Geophys. J.* **95**, 397–419 (1988).
11. Tchavdarov, B., Yarin, A. L. & Radev, S. *J. Fluid Mech.* **253**, 593–615 (1993).

Alzheimer's peptide kills cells of retina *in vivo*

Alzheimer's disease, the commonest form of dementia, is a progressive, age-dependent disorder characterized by the presence of large numbers of senile plaques and neurofibrillary tangles¹. The neuritic plaques consist of an accumulation of amyloid- β peptide ($A\beta$); this substance, which derives from proteolysis of β -amyloid precursor protein (APP), seems to play a central role in the pathology of the disease. The toxicity of $A\beta$ to neurons has already been demonstrated *in vitro*. Here we show that the peptide is also cytotoxic *in vivo*.

Support for the role of $A\beta$ in Alzheimer's comes from reports showing that mutations in the gene encoding APP, and in other genes involved in the disease, lead to overproduction of the more amyloidogenic 42 amino-acid residue long isoform of $A\beta$, $A\beta_{1-42}$. Transgenic mice overexpressing the gene for APP show similar changes, and aggregated $A\beta$ is toxic to cultured neurons of the cerebral cortex or

FELICE FRANKEL

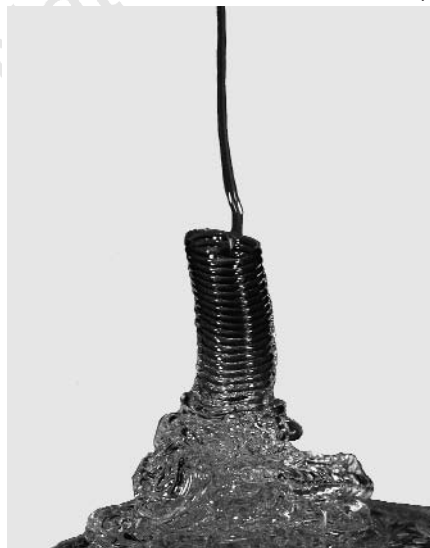


Figure 1 Photograph of a coiling filament of silicone oil (PDMS, $\mu=1.019 \times 10^3 \text{ g cm}^{-1} \text{ s}$, $\rho=1 \text{ g cm}^{-3}$, Gelest), highlighting the constant-diameter coils of diameter of approximately 0.5 cm near the surface.

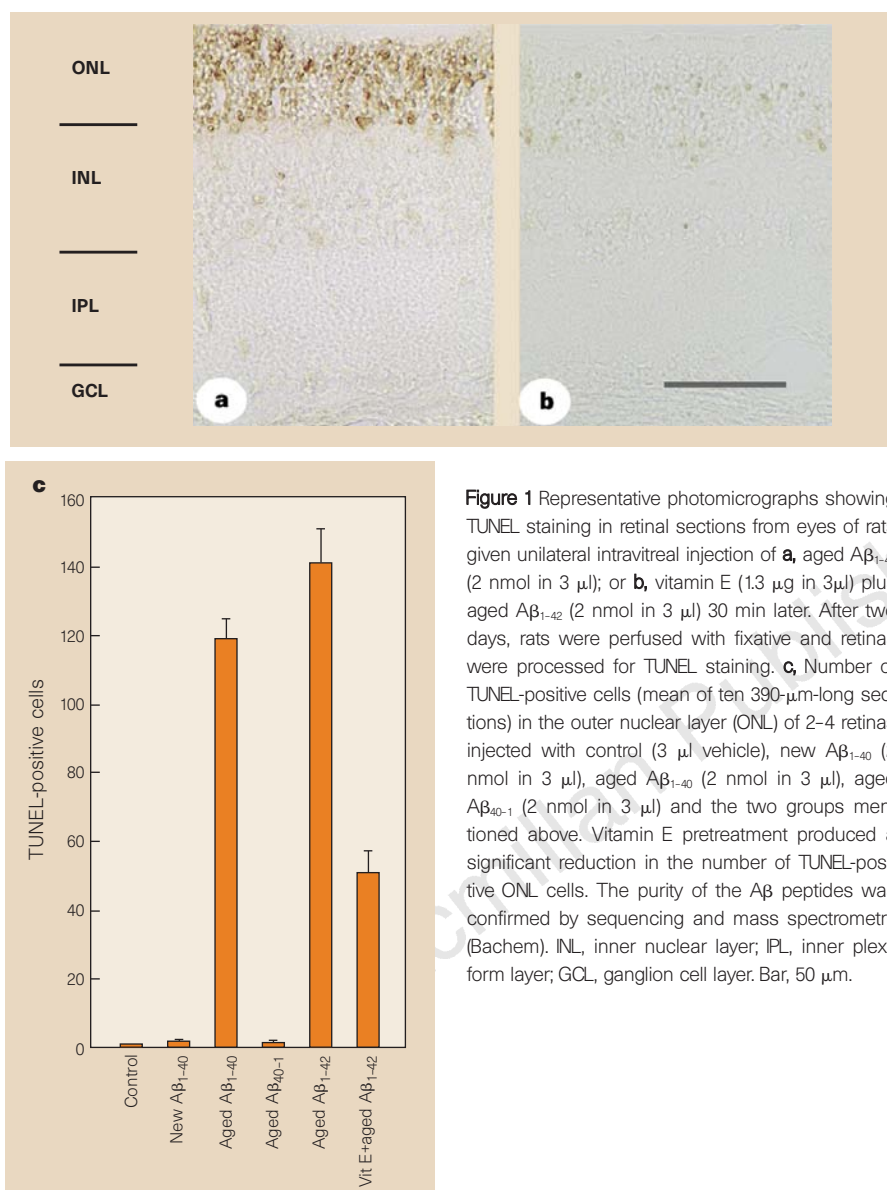


Figure 1 Representative photomicrographs showing TUNEL staining in retinal sections from eyes of rats given unilateral intravitreal injection of **a**, aged $A\beta_{1-42}$ (2 nmol in 3 μ l); or **b**, vitamin E (1.3 μ g in 3 μ l) plus aged $A\beta_{1-42}$ (2 nmol in 3 μ l) 30 min later. After two days, rats were perfused with fixative and retinas were processed for TUNEL staining. **c**, Number of TUNEL-positive cells (mean of ten 390- μ m-long sections) in the outer nuclear layer (ONL) of 2-4 retinas injected with control (3 μ l vehicle), new $A\beta_{1-40}$ (2 nmol in 3 μ l), aged $A\beta_{1-40}$ (2 nmol in 3 μ l), aged $A\beta_{40-1}$ (2 nmol in 3 μ l) and the two groups mentioned above. Vitamin E pretreatment produced a significant reduction in the number of TUNEL-positive ONL cells. The purity of the $A\beta$ peptides was confirmed by sequencing and mass spectrometry (Bachem). INL, inner nuclear layer; IPL, inner plexiform layer; GCL, ganglion cell layer. Bar, 50 μ m.

the hippocampus *in vitro*¹⁻³. Although the mechanisms by which $A\beta$ causes neuronal death are not fully understood, *in vitro* results suggest that an increase in oxidative stress and destabilization of calcium homeostasis^{2,3} are involved.

However, in contrast to these *in vitro* observations²⁻⁴, only a slight neurotoxicity results when $A\beta$ is directly administered to the brain^{5,6}, perhaps because of the rapid efflux of injected compounds from this organ.

In this respect another organ, the retina, offers several advantages. It is an integral part of the central nervous system; its structure is well organized; and, because it is a closed system, injected compounds remain in the vitreous body for a long time⁷.

We investigated the cytotoxic effects of $A\beta$ on retinal cells after intravitreal injection and the amelioration of these effects by the antioxidant vitamin E. The technique used to assess apoptosis and therefore cytotoxicity was terminal deoxynucleotidyl

transferase dUTP-end-labelling (TUNEL) staining. This technique identifies the 3' ends of DNA strands — a phenomenon resulting from the fragmentation of DNA that occurs in apoptotic cells.

When the retinas of rats were injected with either aged $A\beta_{1-42}$ or aged $A\beta_{1-40}$ (aggregated; incubated for 4 days at 37 °C), a distinct band of TUNEL-positive photoreceptor cells was evident in the retina's outer nuclear layer. However, in the inner nuclear layer and ganglion cell layer, only a few scattered cells were TUNEL positive (Fig. 1a). This suggests that proximity to the $A\beta$ injection site is not essential for cytotoxicity.

In addition, all retinas injected with the freshly prepared, 'new' $A\beta_{1-40}$ behaved like vehicle-injected and uninjected retinas in showing only occasional TUNEL-positive cells (Fig. 1c). Therefore the cytotoxic effect was specific to treatment with aggregated $A\beta$. Moreover, unlike active aged $A\beta_{1-40}$, inactive aged reverse $A\beta_{40-1}$ did not induce any changes in photoreceptor cells (Fig. 1c).

No DNA fragmentation was seen in the glial cells. But, as with cultured neural cells *in vitro*⁸, treatment with $A\beta_{1-42}$ or $A\beta_{1-40}$ severely decreased Bcl-2 immunoreactivity in the endfeet and proximal part of radial processes of Müller glial cells.

In vitro studies also suggested that $A\beta$ causes an accumulation of hydrogen peroxide and lipid peroxides in the cells, and that antioxidants protect cells from $A\beta$ toxicity^{2,3}. We found that TUNEL-positive photoreceptor cells in retinas that had been pretreated with vitamin E (1.3 micrograms per 3 microlitres) showed a reduction of about 65% compared with retinas treated with aged $A\beta_{1-42}$ (Fig. 1). This indicates that $A\beta$ cytotoxicity is caused, at least partly, by a free-radical mechanism.

Doubts have been raised in the past about extrapolating *in vitro* findings on $A\beta$ cytotoxicity to the situation *in vivo*. Our results show that some of the proposed mechanisms underlying $A\beta$ neurotoxicity do operate *in vivo*. In addition, the closed 'retina-vitreous' system we describe should serve as an experimental model for further *in vivo* studies and for testing appropriate compounds for their ameliorating effects on $A\beta$ -mediated neurotoxicity.

L. S. Jen, A. J. Hart*, A. Jen, J. B. Relvas, S. M. Gentleman, L. J. Garey, A. J. Patel
MRC Laboratory and Department of Neurodegenerative Disorders, Division of Neuroscience, Imperial College School of Medicine, Charing Cross Hospital, Fulham Palace Road, London W6 8RF, UK
e-mail: a.patel@cxwms.ac.uk

* Present address: The National Hospital for Neurology and Neurosurgery, Queen Square, London, WC1N 3BG, UK

1. Selkoe, D. J. *Annu. Rev. Neurosci.* 17, 489-517 (1994).
2. Mark, R. J., Blanc, E. M. & Mattson, M. P. *Mol. Neurobiol.* 12, 211-224 (1996).
3. Cotman, C. W. & Su, J. H. *Brain Pathol.* 6, 493-506 (1996).
4. Gray, C. W. & Patel, A. J. *Brain Res.* 691, 169-179 (1995).
5. Games, D. et al. *Neurobiol. Aging* 13, 569-576 (1992).
6. Podlinsky, M. B. et al. *Am. J. Pathol.* 142, 17-24 (1993).
7. Mey, J. & Thanos, S. *Brain Res.* 602, 304-317 (1993).
8. Paradis, E., Douillard, H., Koutroumanis, M., Goodyer, C. & LeBlanc, A. J. *Neurosci.* 16, 7533-7539 (1996).

Are retrotransposons long-term hitchhikers?

Transposable elements represent a large fraction of the genomes of eukaryotes, and yet we know little of their origins or stability. Striking examples of cross-species transfer have been discovered among *mariner* elements¹ (transposable elements that are widespread in insects and other animals), confirming the impression that horizontal transfers are essential to the long-term success of transposable elements. We show that R1 and R2, two distantly related non-long-

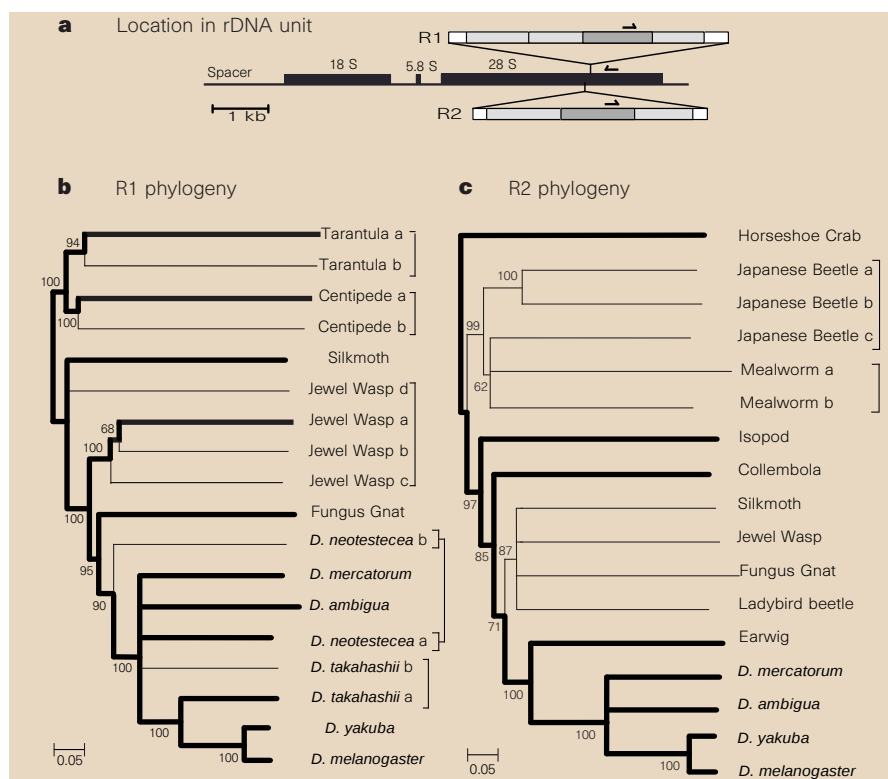


Figure 1 R1 and R2 phylogeny. **a**, Location of R1 and R2 in the rDNA unit. Light shading represents the open reading frames (ORFs); darker shading represents the reverse transcriptase (RT) domain. Short arrows: PCR primer positions^{3,5,7} used to amplify the 3' half of each element. **b**, **c**, Phylogeny of R1 and R2 elements. Shown are the 60% consensus trees based on 'neighbour-joining' methods with bootstrap values indicated¹⁰. Trees have been rooted by other non-long-terminal-repeat retrotransposons using their reverse transcriptase domains. Brackets to the right indicate multiple families in the same species, with each family given a letter classification. Bold lines represent a lineage consistent with the species phylogeny. The topology of each tree is also supported by 'maximum parsimony' algorithms and 'maximum likelihood' methods based on both amino-acid and nucleotide sequences, except for minor rearrangement of the multiple R2 lineages in Japanese beetle and mealworm. Accession numbers for previously published elements can be found in refs 3–7; those for the newly determined elements are AF015489 and AF015813–AF015821.

terminal-repeat retrotransposons which insert at specific sites 74 base pairs apart in 28S ribosomal RNA genes², have been maintained by vertical transmission since the origin of the phylum Arthropoda, that is, for at least 500 million years.

The insertion specificity of R1 and R2 (Fig. 1a) offers advantages for assessing the extent of these retrotransposons' vertical transmission in an old and diverse lineage, because defective elements are eliminated from the host, and all elements can be cloned irrespective of their relationship to previously characterized elements. In previous studies^{3–6} we have shown that R1 and R2 are stable components of the ribosomal DNA loci within the genus *Drosophila*, and are evolving at similar rates to typical nuclear genes.

We cloned the 3' half of R2 elements from the insects: *Tenebrio molitor* (mealworm), *Hippodamia convergens* (ladybird beetle), *Forficula auricularia* (earwig), *Anurida maritima* (collembola), and the non-insect arthropods: *Porcellio scaber* (isopod) and *Limulus polyphemus* (horseshoe crab). The 3' half of R1 elements were

cloned from *Dugesiella* sp. (tarantula) and *Scolopendra* sp. (centipede). The sequences were obtained by the polymerase chain reaction using one primer to the YADD region of the reverse transcriptase domain, and a second primer to the 28S rRNA gene downstream of the insertion sites^{3–5}.

These sequences were combined with the sequences of elements previously obtained from the insects *Sciara coprophila* (fungus gnat), *Bombyx mori* (silkworm), *Nasonia vitripennis* (jewel wasp) and *Popillia japonica* (Japanese beetle)⁷. Multiple lineages of either R1 or R2 elements were found in several of these species. As found previously^{5,7}, the level of nucleotide divergence between elements is either less than 1% (the same family) or greater than 25% (different families).

To resolve the deep phylogenies of these R1 and R2 elements, conceptual translations of their open reading frames (minimum 360 and 480 amino-acid positions for R1 and R2, respectively) were used in phylogenetic reconstructions (Fig. 1). The phylogenies determined for R1 reveal that in each arthropod, at least one family of ele-

ments is congruent with the generally accepted phylogeny of the host species (lineages represented by bold lines), indicating stable vertical inheritance of the elements. Among the species with multiple R1 families, there are several examples where one of the families appears to be more divergent than expected (*D. takahashii*, b; *D. neotestacea*, b; jewel wasp, d). The R2 phylogeny (Fig. 1c) differs from the R1 phylogeny in that multiple families are only found in beetles.

There are two apparent violations of the host species phylogeny on the R2 tree: the families from Japanese beetle and mealworm are too divergent, branching before the other insects; and the *Drosophila* R2 elements are more similar to those of an earwig than to another dipteran (fungus gnat). Although these incongruities could be due to old horizontal transfers, as they represent older than expected divergences, it is more likely that different active lineages of R2 have been propagated in various insects.

This analysis of R1 and R2 sequences suggests these elements have been present during the entire history of the arthropod lineage, a time that is estimated to be at least 500 million years^{8,9}. Because their insertion precludes the use of the 28S rRNA gene, R1 and R2 are presumed to be deleterious. How then do we account for their remarkable stability? Is it a case of selfish elements occupying an ideal niche in a conserved multigene family, or do these elements provide useful functions for their hosts which have for some reason maintained them as transposable elements?

Although these phylogenetic data cannot answer these questions, they do establish the potential for retrotransposons to be long-term hitchhikers of eukaryotic genomes. Indeed, if R1 and R2 are used as standards for divergence time, the lineages of many non-long-terminal-repeat retrotransposons appear to date back to the Cambrian era.

William D. Burke, Harmit S. Malik, Warren C. Lathe III, Thomas H. Eickbush
Department of Biology, University of Rochester,
Rochester, New York 14627-0211, USA
e-mail: eick@uhura.cc.rochester.edu

- Robertson, H. M. *Nature* 362, 241–245 (1993).
- Jakubczak, J. L., Burke, W. D. & Eickbush, T. H. *Proc. Natl. Acad. Sci. USA* 88, 3295–3299 (1991).
- Eickbush, D. G. & Eickbush, T. H. *Genetics* 139, 671–684 (1995).
- Eickbush, D. G., Lathe, W. C. III, Francino, M. P. & Eickbush, T. H. *Genetics* 139, 685–695 (1995).
- Lathe, W. C. III, Burke, W. D., Eickbush, D. G. & Eickbush, T. H. *Mol. Biol. Evol.* 12, 1094–1105 (1995).
- Lathe, W. C. III & Eickbush, T. H. *Mol. Biol. Evol.* 14, 1232–1241 (1997).
- Burke, W. D., Eickbush, D. G., Xiong, Y., Jakubczak, J. L. & Eickbush, T. H. *Mol. Biol. Evol.* 10, 163–185 (1993).
- Bowring, S. A. *et al. Science* 261, 1293–1298 (1993).
- Wray, G. A., Levinson, J. S. & Shapiro, L. H. *Science* 274, 568–573 (1996).
- Thompson, J. D., Higgins, D. G., Gibson, T. J. *Nucleic Acids Res.* 22, 4673–4680 (1994).

Twin paradoxes of the space age

Countdown: A History of Spaceflight

by T. A. Heppenheimer
Wiley: 1997. Pp. 398. £24.95, \$30

Korolev: How One Man Masterminded the Soviet Drive to Beat America to the Moon

by James Harford
Wiley: 1997. Pp. 392. £24.95, \$30

Something New Under the Sun: Satellites and the Beginning of the Space Age

by Helen Gavaghan
Copernicus: 1998. Pp. 300. \$26, £15

Space and the American Imagination

by Howard E. McCurdy
Smithsonian Institution Press: 1997. Pp. 294.
\$29.95, £19.95

Alex Roland

My generation of Americans, born during the Second World War, recalls the first landing on the Moon in 1969 as the defining moment in the space age. Today's students think of the Challenger accident, or at least that was the icon most familiar to the class I lectured in autumn 1997. Some could recall the rescue of the Hubble space telescope; a few mentioned the previous summer's charming performance of the Sojourner spacecraft on Mars. But none of them knew or cared very much about the remarkable history of scientific and technological achievement conveyed in these books.

What happened to the space age? Do our callow youth simply take for granted the marvels wrought by their parents and grandparents? Or has spaceflight lost its way and its identity? Do many people know or care that an international space station will start tak-

ing shape in orbit later this year? Can any of those people explain what that station is supposed to be or do?

These books do not answer that question directly. They do, however, shed light on two historical paradoxes that may provide an answer. First, there is a paradox of purpose that juxtaposes military and civilian goals in space. Spaceflight as we know it is an artefact of the Cold War, the end of which robbed spaceflight of its engine and its compass. Second, a paradox of method pits manned spacecraft against unmanned ones. Manned spaceflight dominated public interest in the Cold War, yet automated spacecraft did most of the work. Each of these books contributes to an understanding of how these paradoxes arose.

T. A. Heppenheimer, an associate fellow of the American Institute of Aeronautics and Astronautics (AIAA), is the author of six previous books and many articles on aviation and space. *Countdown* tells the history of spaceflight in three acts. The first third of the book focuses on the traditional cast of pioneers — Konstantin Tsiolkowski, Hermann Oberth, Robert Goddard, Wernher von Braun, Theodore von Kármán and a host of lesser names. To these he adds two who have received less attention in the literature. Sergei Korolev, the subject of James Harford's biography, carries the burden of the Soviet story; he is often depicted as the Soviet von Braun. William Bollay appears as the American genius behind the family of rockets that powered everything from the aborted Navaho missile of the 1950s to the Saturn rocket that launched the Apollo missions of the 1960s and 1970s. In a sharp departure from conventional wisdom, Heppenheimer calls Bollay "America's top post-war rocket leader", an accolade usually reserved for von Braun.

The middle third of Heppenheimer's book recreates the golden age of spaceflight, from Sputnik 1 in 1957 to the first manned Moon landing in 1969. The Cold War drove the space race and energized its history. The visionaries realized their dreams and infected the world with their enthusiasm and faith. The US president at the time, John Kennedy, called space "the new

ocean", suggesting that the Moon mission would inaugurate a new era in history.

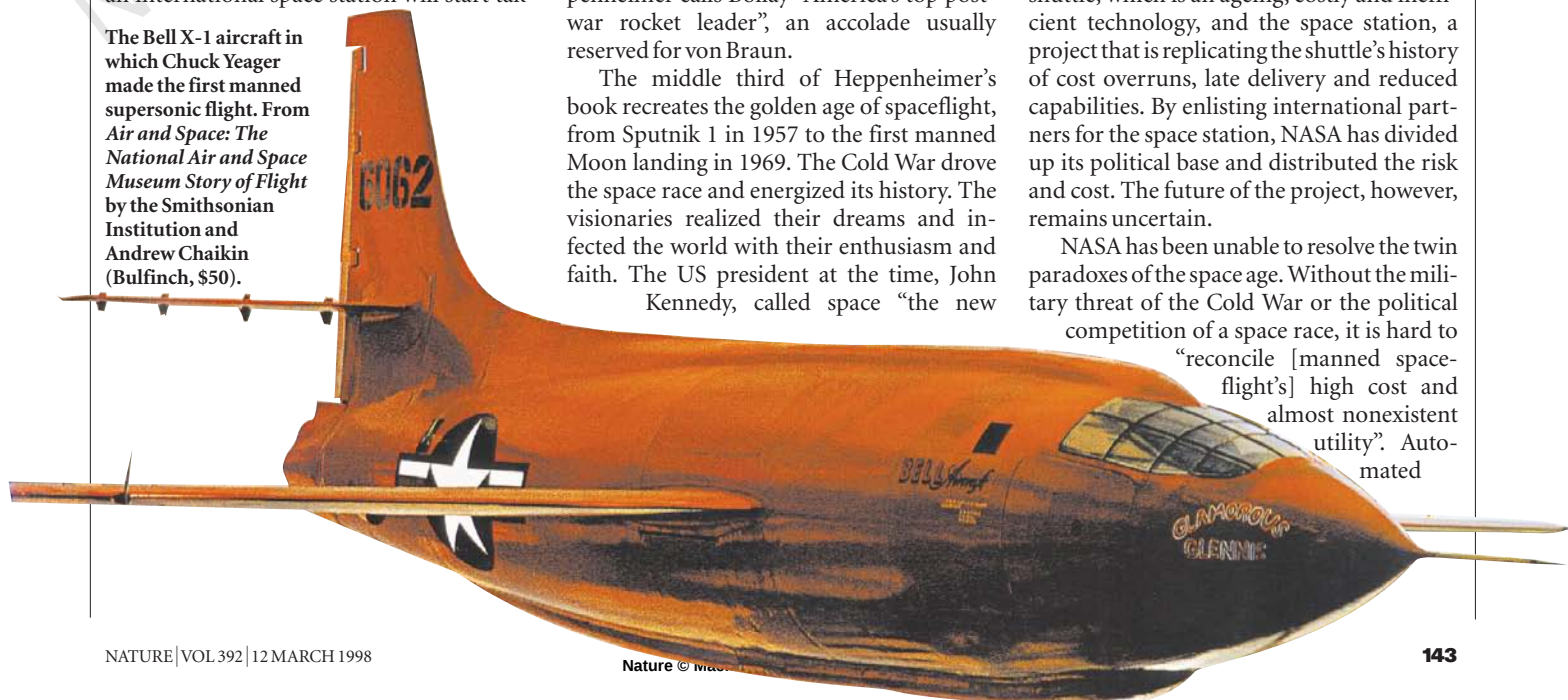
The final act of Heppenheimer's play, however, describes a "retreat from the moon". "The space program", he says, "[had] to find its role in a prosaic world where the glow of Kennedy's challenge had faded." The Soviets lost Korolev to cancer during the Moon race and fell increasingly behind the United States. They turned their attention to endurance records on space stations. The United States, meanwhile, put its future in space on the space shuttle, a launch vehicle flawed in both concept and execution. While it was under development in the 1970s, "unmanned spacecraft were carrying out the real work of space".

Heppenheimer is unable to integrate the stories of manned and unmanned spacecraft. His narrative stops in the third act for an isolated chapter on the automated spacecraft that fly to the planets, orbit the Earth, conduct scientific investigations, transmit communications, analyse the weather, scrutinize the environment, and carry on a huge range of other useful functions. This catalogue of utility completed, Heppenheimer inevitably returns to Challenger in January 1986. When the US Air Force lost a Titan III rocket just three months later, one space programme insider lamented "we are essentially out of business".

With the Cold War over, Russia in receivership, and other countries shying away from their own programmes of manned spaceflight, the United States finds itself racing alone. Daniel Goldin, administrator at the US space agency NASA, has refocused the agency on faster, smaller, cheaper projects. But he is burdened with the space shuttle, which is an ageing, costly and inefficient technology, and the space station, a project that is replicating the shuttle's history of cost overruns, late delivery and reduced capabilities. By enlisting international partners for the space station, NASA has divided up its political base and distributed the risk and cost. The future of the project, however, remains uncertain.

NASA has been unable to resolve the twin paradoxes of the space age. Without the military threat of the Cold War or the political competition of a space race, it is hard to "reconcile [manned spaceflight's] high cost and almost nonexistent utility". Automated

The Bell X-1 aircraft in which Chuck Yeager made the first manned supersonic flight. From *Air and Space: The National Air and Space Museum Story of Flight* by the Smithsonian Institution and Andrew Chaikin (Bulfinch, \$50).



spacecraft deliver the real pay-off in space, but they are invisible to the public. But the history of spaceflight is unfinished. So too is Heppenheimer's book; the publisher left the last page of his text blank, inadvertently one assumes. Authors sue for such incompetence, but in this case it might be a fitting conclusion to this useful but necessarily inconclusive study.

History gave James Harford a much more satisfactory ending to his biography of Korolev. The genius behind the S-7 family of Soviet launch vehicles survived internal exile in the gulags during Stalin's purges of 1938 to design aircraft for the motherland during the Second World War and achieve complete rehabilitation and ultimate canonization in the race to build intercontinental ballistic missiles and space launch vehicles. Like von Braun, Korolev dreamed of spaceflight and made common cause with a totalitarian dictator to get the state resources necessary to pursue his dreams. He designed and built the rocket that launched Sputnik. Then he tried frantically to stay ahead of the United States in the space race he had started. By the time he died on the operating table in 1966 of complications from cancer, the United States was ahead. Korolev presided over the glory years, though he surely knew at the end that the race was lost. He did not live to know that his dream of human space exploration would atrophy as well.

Harford, executive director emeritus of the AIAA, has produced a competent, earnest biography of Korolev, based mostly on secondary sources, that does not rise much above derivative eulogy. Nor does it help very much our understanding the larger dilemmas that surround space activity at the end of the twentieth century.

Helen Gavaghan, a science writer and editor, tells an even narrower story, but on a higher plane. She operates in the realm to which Heppenheimer devoted only a chapter: automated spacecraft. In fact, she covers only the first decade of automated spacecraft, limiting herself further to navigation, weather and communication satellites in near-Earth orbit. The book is therefore rich in technical detail, some of it new to the published literature.

She tells these stories both clearly and fully, in keeping with the goals of the Sloan Technology Series, which underwrote her project. Heroes drive the developments she describes, overcoming the scepticism of bureaucrats and engineers to see their concepts reified in working spacecraft. William H. Guier and George C. Weiffenbach, for example, proved in the wake of Sputnik that they could track satellites by measuring their Doppler shift. Then they demonstrated how the same technique could be reversed to provide passive navigation signals to military ships and aircraft. Thus was born the US Transit worldwide navigation system, a fore-



Stealth fighter: the Lockheed F-117A Nighthawk bomber aircraft has the ability to evade enemy radar and is controlled in flight by sophisticated computers. From *Air and Space* (see caption on page 143).

runner of today's global positioning system.

Gavaghan's book pulses with the interaction between military and civilian technology. Dwight Eisenhower, then US president, approved the satellite project of the international geophysical year (1957–1958) as a means of establishing the right of satellite overflight. He wanted the precedent to legitimate reconnaissance satellites. Once the principle was established, however, civilian applications gained independent support. Although navigation remained closely linked to military applications, weather and communications were soon decoupled from pilot military programmes.

The marvels of global positioning technology, weather satellites and worldwide communication were achieved entirely by automated spacecraft, both military and civilian. As Heppenheimer makes clear, rapid advances in microelectronics allowed these unmanned spacecraft to do what Korolev, von Braun and their contemporaries thought human crews would be required to do *in situ*. But only communications satellites have paid their own way; all other space activity remains, as Heppenheimer puts it, the "sport of governments".

Howard McCurdy, professor of public affairs at American University in Washington DC and an acknowledged authority on the US space programme, comes closest to explaining why the term 'space age' resonates with his generation and mine but holds no salience for young people. The answer lies in unrealistic expectations. At first sceptical of 'Buck Rogers' proposals, the US public were converted in the 1950s by a band of space propagandists, led by the greatest of all salesmen, von Braun. Primed by the success of aviation and advances in consumer technology, Americans conflated science fiction

with science fact, and romance with reality. In McCurdy's terms, they let imagination rule reason. They subscribed to an agenda beyond the reach of the treasure they were willing to invest. At the heart of that agenda was man (not woman) in space. Unable to imagine what space might really be like, they measured it in terms of their own history. It was the new world waiting for Christopher Columbus. It was the frontier awaiting pioneers. For a while the space station was a frontier fort, an outpost of civilization, a station on a heroic voyage of discovery.

In reality, of course, it was none of these. By attempting to fashion the reality to suit their history, Americans failed to distinguish between the past and the future. The future they imagined failed to materialize, their dreams have evaporated, their fund of optimism is bankrupt, and their children neither believe nor care. McCurdy tells a richly detailed and persuasively balanced story that explains the paradox of manned spaceflight, at least in the US context.

The paradox of the military and civilian realms looms large in the works of Heppenheimer, Gavaghan and Harford. They make clear that the space age was born of military technology and driven in the first decades by the imperatives of the Cold War. But they do not make clear the curious nexus between this historical phenomenon and the paradox of manned against unmanned spaceflight.

Even though the early astronauts and cosmonauts were military officers, even though they first went aloft on modified ballistic missiles, even though early satellites were developed with military purposes in mind, and even though men in space were expected to play largely military roles, the actual course of military activity in space has been almost entirely unmanned. Satellites

used for reconnaissance, communications, weather and navigation have quietly gone about their business without human operators, contributing to the peaceful resolution of the Cold War and the maintenance of 'the long peace'. The public is confused and indifferent to space because most of the important work there, both military and civilian, has been done out of public view by unmanned spacecraft. The manned spacecraft, as Heppenheimer observes, have received most of the attention but have produced few returns.

The visionaries who dreamed of manned space exploration would argue that humans will eventually travel to the planets and beyond. Perhaps they will. But these books show that they will not do it with Cold War technology, of which the international space station is the latest incarnation. My students will have to invent their own generation of spacecraft for that purpose. But right now, they do not seem all that interested. □

Alex Roland is in the Department of History, Duke University, Durham, North Carolina 27708, USA.

Site reading

Monte Verde – A Late Pleistocene Settlement in Chile, Vol. 2: The Archaeological Context and Interpretation

by Tom D. Dillehay

Smithsonian Institution Press: 1997.

Pp. 1,071. \$155, £120.95

Nicholas J. Saunders

Of all the issues in American archaeology, few have been so bitterly contested as that concerning the origins of the continent's first human settlers. There is agreement that the first Americans came from Siberia across the Beringian land bridge that linked the Eurasian and American landmasses, but the dating of this episode has split the archaeological community for decades. Until now, there have been two schools of thought: that championing the Clovis people around 11,000 years ago, and another that favoured much earlier dates, anywhere between 13,000 and 30,000 years ago.

The Clovis occupation, with its clearly artefactual fluted points and securely dated contexts, was recognized long ago. Those who advocated a pre-Clovis human presence saw their discoveries undermined by accusations of controversial dating, problematic stratigraphy, wishful thinking and endless debates about whether or not chipped stones and bones were manmade tools or naturally occurring 'eco-facts'.

Into this cockpit of competing egos and academic reputations came Tom D. Dillehay and the Chilean site of Monte Verde in 1977. From the start, Monte Verde was excitingly different. Unlike previous putatively early

sites, Monte Verde was not a rock shelter but an open-air location by a creek. Its cultural remains were not confined to ambiguous stones but included eminently datable wooden artefacts, clay-lined firepits and the remains of meat, hide and almost 70 plant species. This was a real campsite, home to a community of hunter-gatherers who exploited a wide range of food resources, including wild potato and plant species known ethnographically for their medicinal properties.

The implications were clear. If, at the southern tip of South America, Monte Verde had been occupied by around 12,500 years ago — as a clutch of radiocarbon dates indicated — then humans must have entered the Americas well before Clovis times. And there was more: Dillehay also discovered an even earlier site nearby, buried in a promontory of an old lagoon, that yielded dates of approximately 33,000 years ago.

Reports on aspects of the site's remains and palaeo-environmental features have appeared since 1977, and have been attacked or ignored by Dillehay's opponents. Then, in early 1997, Dillehay took a calculated risk by flying a group of leading sceptics to Chile and letting them examine the site, its remains and his painstaking work at first hand. By most accounts, they were heavily persuaded that Monte Verde was indeed indisputably of pre-Clovis age.

This long preamble is necessary to appre-

ciate the continent-wide significance of both the site and the book, the latter being a watershed publication in studies of early American prehistory. In a book organized into 22 chapters and 16 appendices, and spread across more than 1,000 pages, Dillehay has put together what must be one of the most persuasive and impressive interdisciplinary volumes on any kind of archaeology for a generation.

As many specialists have long acknowledged, breaching the pre-Clovis barrier was never going to be easy. It was as much a psychological as an archaeological divide. In this sense, the sceptics did Dillehay a favour: by demanding ever greater detail and rigour they threw down a challenge that few would have taken up, let alone successfully overcome.

Although more than 70 scientists have contributed to this book, Dillehay has authored or co-authored most of the chapters, providing a coherence and fluency often missing from such monographs. Critical issues are all addressed in cogently argued and well illustrated detail. Research design and method, context, taphonomic analyses of wood assemblages, micro-use-wear of lithic assemblages, animal and plant remains, microtopographical features of use surfaces, and social organization all have their own chapters, and are buttressed by appendices written by specialists in everything from wood conservation to stone

Of tombs and temples



"The greatest pyramid-builder in Egyptian history" was King Sneferu, who came to the throne in around 2575 BC. His four pyramids included the Bent pyramid at Dahshur (above) and used a total mass of stone that exceeds even that of the Great Pyramid at Giza, built by his

son Khufu. From *The Complete Pyramids* by Mark Lehner (Thames and Hudson, £24.95, \$39.95). The author examines practical aspects of the monuments' construction — the quarries, ramps and tools used — as well as their associated cosmology and iconography.

petrology, microscopy and tree-ring studies.

Every step of the way Dillehay has anticipated possible objections and been careful to explain, describe and evaluate not only his methodology but also the smallest category of potential evidence. For example, no less than 1,461 pieces of charcoal were examined for spatial distribution, degree of charring and identification, where possible, of wood taxa. This level of recording yielded valuable insights into the selection of specific kinds of wood for fuel, and the use of fire for cooking, heating and tool manufacture.

From such detail comes the realization that Monte Verde is one of those rare books that, as an impressive and influential scholarly achievement, set a benchmark for all future work. □

Nicholas J. Saunders is in the Department of Anthropology, University College London, Gower Street, London WC1E 6BT, and the Department of Archaeology, University of Southampton, Highfield, Southampton SO17 1BJ, UK.

The trouble with sex

Sex on the Brain: The Biological Differences Between Men and Women

by Deborah Blum
Viking: 1997. Pp. 329. \$24.95

Melissa Hines

It is hard to write a good book about psychological sex differences. The topic is relevant to attitudes and social policy in many areas, including who should take primary responsibility for childcare, whether education should differ for girls and boys, whether laws should prohibit certain sexual practices, and to what extent choice of occupation is dictated by sex.

The scope of this book is broad. For a general idea of its factual accuracy, I looked at its coverage of my own research areas: sex differences in the human brain and the role of androgen and oestrogen in shaping gender development. Here the disappointment began.

A simple example is whether prenatal exposure to high concentrations of androgen ('male' hormone) improves mathematical ability. In this book Deborah Blum says that it does, suggesting that prenatal brain programming by hormones causes males to excel at mathematics. But the only two existing studies of the relationship between prenatal androgen and mathematical ability found impaired ability.

The trouble with sex is that people, scientists and non-scientists alike, have their own theories about how men and women differ. Developmental psychologists call these theories 'gender schemas' and find that they influence memory. For example, children shown pictures of male nurses and female

doctors later confidently recall seeing female nurses and male doctors. Those writing popular books on psychological sex differences often seem to have similar problems with remembering research findings.

The book relies heavily on theories from evolutionary psychology. These theories are difficult to test empirically and are particularly prone to contamination by gender schemas. They seem sensible only because they correspond to popular preconceptions. For instance, Blum describes a theory that men prefer blondes because men are evolutionarily predisposed to prefer youthful mates, and blondness is a sign of youth. But the blondness of youth usually disappears by puberty.

If preference for blondes reflects an evolutionary predisposition to youthful mates, men would be predisposed to paedophilia. If this suggestion seems more absurd than the original, it is because prevailing gender schemas suggest that men have innate preferences for young women but not children.

The book also suffers from a failure to screen information for reliability. Everything, from replication of results in refereed journals, through presentations and comments at scientific conferences, to reports in the popular press, is given equal credibility. This casual approach to facts, combined with phrases like "those nitpicky cell-counting studies", left me feeling as if I had eaten too much junk food.

Some might argue that accuracy is unimportant, because a popular book can stimulate interest in a research area. But facts do matter. Take again the supposed androgen-induced predisposition to mathematical ability. Research into attitudes and performance shows that an assumption of inability can become a self-fulfilling prophecy. This misinformation itself could therefore impair mathematical performance in girls and women.

Research into psychological sexual differentiation has produced surprising findings that sometimes contradict prevailing gender schemas. For instance, although we think of sex as being determined by the sex chromosomes (XX or XY), their role is minor compared with that of gonadal hormones. Also, although hormones from the male gonads are essential for masculine development, hormones from the female gonads have relatively little influence on feminine psychological development. Even more remarkably, what we think of as the major 'female' hormone, oestrogen, can have powerful masculinizing influences during development. For further information consult scientific publications, but mind your gender schemas. □

Melissa Hines is in the Department of Psychology, City University, Northampton Square, London EC1V 0HB, UK.

New in paperback

One River: Science, Adventure and Hallucinogenics in the Amazon Basin

by Wade Davis
Simon and Schuster, £7.99
 Reviewed in *Nature* 384, 229 (1996)

The Conscious Mind

by David J. Chalmers
Oxford University Press, £9.99
 Reviewed in *Nature* 381, 123 (1996)

Humphry Davy: Science and Power

by David Knight
Cambridge University Press, £16.95, \$27.95
 Reviewed in *Nature* 359, 785 (1992)

Infinite Potential: The Life and Times of David Bohm

by F. David Peat
Helix, \$16
 Reviewed in *Nature* 385, 592 (1997)

The End of the World: The Science and Ethics of Human Extinction

by John Leslie
Routledge, £9.99
 Reviewed in *Nature* 380, 296 (1996); see also *Nature* 387, 338 (1997)

Catastrophism: Asteroids, Comets, and Other Dynamic Events in Earth History

by Richard Huggett
Verso, £14, \$19
 Original 1990 edition reviewed in *Nature* 345, 778 (1990)

Bats: Biology and Behaviour

by John D. Altringham
Oxford University Press, £17.99

Corrections

- The full list of co-authors of *Principles of Development* by Lewis Wolpert (reviewed in *Nature* 391, 857; 1998) should have read: Rosa Beddington, Jeremy Brockes, Thomas Jessell, Peter Lawrence and Elliot Meyerowitz.
- Some lines dropped off Derek Fordham's review of *Nansen* (*Nature* 391, 137; 1998). The sixth paragraph should start: "When it became clear the *Fram* would not drift over the North Pole, Nansen and Hjalmar Johansen left the ice-locked vessel at latitude 84° 4' and attempted to reach the pole over the ice. They turned back at latitude 86° 14', having established a new farthest-north record..."
- In the review of *Next of Kin* (*Nature* 390, 246; 1997), an editorial change rendered the fictitious veterinarian "Dr Doolittle" (John) as "Dr Doolittle" (Eliza).

Lattice effects in magnetoresistive manganese perovskites

A. J. Millis

The discovery of spectacularly large magnetoresistive responses in a class of metallic manganese oxides has raised hopes that these compounds might be of practical utility. But regardless of whether this promise is realized, these materials provide an ideal system in which to elucidate the properties of metals in which electron–lattice interactions play a key role.

Transition-metal oxides have long been the subject of study, because they exhibit a wide range of exotic and still imperfectly understood structural, magnetic and electronic behaviour. This behaviour cannot be explained within the context of the usual one-electron band theory that accounts well for the properties of most other solids, indicating the importance of strong electron–electron and electron–lattice correlations. The properties of these materials continue to surprise; for example, 1986 saw the discovery of high-temperature superconductivity in materials based on copper oxide.

More recently, attention has become focused on a certain class of manganese oxides, the manganite perovskites. Although these materials have been studied for many years¹, the current burst of activity was stimulated by reports by Helmholtz *et al.*² and Jin *et al.*³ of spectacularly large—“colossal” as Jin *et al.*³ put it—magnetoresistance in this family of compounds. Magnetoresistance, the variation of electrical resistance with magnetic field, is crucial to several areas of technology, such as magnetic data storage, and much of the impetus for the present interest in the manganites stems from the possible utility of their magnetoresistive properties. Whether the manganite perovskites currently under consideration will prove technologically useful is still far from clear. But the observation of colossal magnetoresistance has stimulated a considerable amount of work aimed mainly at understanding and improving their magnetoresistive properties, and at examining other related classes of transition-metal oxides (such as spinels, pyrochlores and magnetites) which display similar behaviour and may have more technologically convenient properties.

But here I will argue that the manganites are also important for basic condensed-matter physics for quite a different reason. In these materials, the interaction between the electrons and lattice vibrations (phonons) is unusually strong, leading to a wide range of striking physical phenomena and, most crucially, can be ‘tuned’ over a wide range by variation of chemical composition, temperature and magnetic field. These materials therefore provide an unprecedented opportunity to study the poorly understood physics of systems in which a high density of electrons is strongly coupled to phonons and, in particular, to elucidate the interplay between local structural deformations and global properties. This interplay is becoming the focus of attention in many contexts, including conducting polymers⁴ and ferroelectrics⁵ (for a summary of other recent work on local structures, see ref. 6).

Basic phenomenology

The ‘colossal magnetoresistance’ (CMR) manganites are compounds based on the ABO_3 perovskite structure shown in Fig. 1a. The most widely studied family has the chemical formula $Re_{1-x}A_xMnO_3$ (where Re is a rare earth such as La or Nd, and A is a divalent alkali such as Sr or Ca), but there is also growing interest in the ‘Ruddlesden–Popper’ series of layered materials $Re_{n+1}Mn_nO_{3n+1}$. The important electrons are the Mn *d* electrons, of

which there are $4 - x$ in $Re_{1-x}A_xMnO_3$. The atomic physics of manganese is such that all of the electrons on a given Mn ion in these compounds must have parallel spin, so each Mn ion has a magnetic moment.

Changing the carrier concentration x produces a variety of phases, which may be characterized by their magnetic, transport and ‘charge-ordering’ properties. A phase diagram in the doping (x) and temperature (T) plane for one representative material, $La_{1-x}Ca_xMnO_3$, is shown in Fig. 2. At $x = 0$, the material is ‘insulating’ (that is, it has a resistivity which is very high and grows as the temperature is lowered) at all temperatures. The material is paramagnetic (it has no long-range magnetic order) at high temperature, but below about 140 K it becomes a $(0, 0, \pi)$

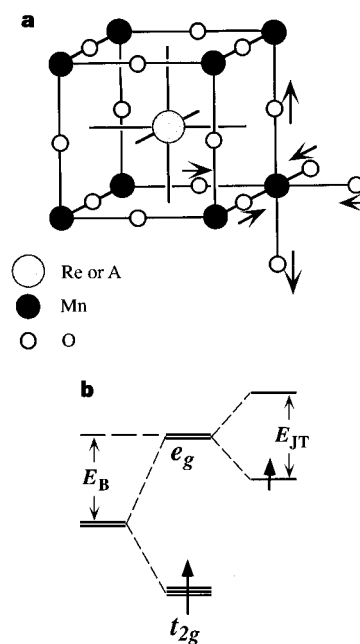


Figure 1 Crystal structure and electronic structure of the manganite perovskite. **a**, The basic perovskite structure, with oxygen motions in one Jahn–Teller distortion indicated by arrows. **b**, Schematic energy levels of the Mn ion. The central portion is for the undistorted lattice; it shows the low-lying t_{2g} levels occupied by three parallel-spin electrons forming a $S_e = 3/2$ core spin (indicated by large arrow), and the higher-lying e_g doublet (two-fold degenerate in the ideal perovskite crystal structure). The right-hand portion of the figure shows that if the e_g level is singly occupied, a Jahn–Teller distortion of the surrounding O_6 octahedron may occur (see **a**), which would split the e_g doublet by an energy E_{JT} . The left-hand portion of the figure shows that if the e_g level is unoccupied, a ‘breathing’ distortion may occur (see text) which lowers the energy of the unoccupied e_g doublet by an amount E_B relative to its energy in the ideal structure.

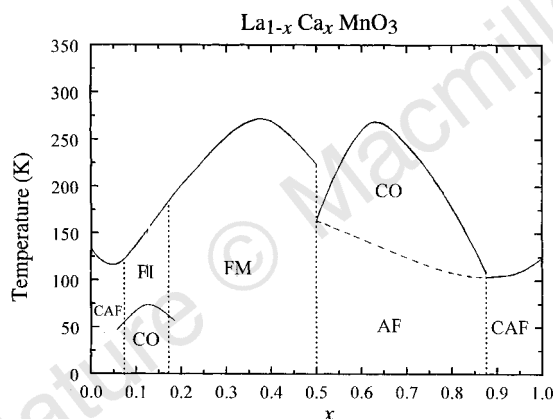
antiferromagnet: in a given plane perpendicular to the crystal c -axis the magnetic moments on Mn sites are aligned, but the moment direction alternates from plane to plane. The ground state remains insulating for $x < x_{\text{MI}} \approx 0.2$ but the magnetic order changes in a complicated (and still controversial) way, eventually becoming a fully polarized ferromagnet (all spins aligned in the same direction). In the phase diagram of Fig. 2, this sequence of phases is denoted by CAF (for canted antiferromagnet) and FI for ferromagnetic insulator. One sees also that charge order (CO—a periodic pattern of Mn sites in different valence states) develops inside the ferromagnetic insulator phase. At $x = x_{\text{MI}}$ the ground state changes from insulating to metallic. This is shown in Fig. 2 as the vertical dotted line separating the FI (ferromagnetic insulator) and FM (ferromagnetic metal) phases. For $x_{\text{MI}} < x < 0.5$ the ground state is a ferromagnetic metal. For $x > 0.5$ the ground state again becomes insulating and antiferromagnetic, and is in addition ‘charge-ordered’.

As the temperature is raised in the region $x_{\text{MI}} < x < 0.5$, there is a ferromagnet–paramagnet transition at a $T_c(x) \approx 200$ –250 K. The magnetic phase boundary also separates a low-temperature ‘metallic’ phase from a high-temperature ‘insulating’ phase. (Here ‘metallic’ and ‘insulating’ have quotation marks because a precise distinction between metal—finite resistivity—and insulator—infinite resistivity—is only possible at zero temperature; here the terms are used in the qualitative sense that the resistivity at $T < T_c(x)$ is relatively low and increases as T is increased, whereas at $T > T_c(x)$ the resistivity is very high and (for most x) decreases as T is increased). The magnetoresistance for $T \approx T_c(x)$ can be very

large: some representative data⁷ are shown in Fig. 3a. The large magnetoresistance is perhaps not surprising: ferromagnetic transitions are in general very sensitive to applied magnetic fields and in the materials of interest the ferromagnetic transition is also a ‘metal’–‘insulator’ transition, so it is natural to expect the resistivity to depend strongly on magnetic field in the vicinity of this transition. What is surprising is the very existence of the metal–insulator transition: it implies there is a mechanism for localizing the carriers at $T > T_c$ and ‘turning off’ the localization as T is lowered through T_c .

Electronic and magnetic structure

In the ideal perovskite structure, each Mn is in a locally cubic environment and the crystal field splits the five d orbitals into a t_{2g} triplet and an e_g doublet. These levels are shown in the central portion of Fig. 1b. The t_{2g} levels seem to lie substantially (~ 2 –4 eV) below the e_g levels. The on-site Coulomb repulsion is apparently strong enough that (1) no d orbital may be occupied by more than one electron and (2) all electron spins on a given Mn are ferromagnetically aligned by a large Hund’s-rule coupling. The resulting physical picture is thus that 3 of the $(4 - x)$ d electrons fill up the t_{2g} levels, forming an electrically inert core spin S_c of magnitude $3/2$, while the remaining $(1 - x)$ electron goes into a linear combination of e_g orbitals and may move through the crystal subject to the constraint that when it is on site i its spin must be parallel to S_c . This means that the amplitude for a carrier to hop from site i to site j is modulated by the overlap between having its spin parallel to S_c and



◀ **Figure 2** Phase diagram of $\text{La}_{1-x}\text{Ca}_x\text{MnO}_3$ showing magnetic and structural phase boundaries (S.-W. Cheong, personal communication). The horizontal axis shows the value of x in the formula. Phases include charge-ordered (CO), antiferromagnet (AF), canted antiferromagnet (CAF), ferromagnetic metal (FM), ferromagnetic insulator (FI). The unlabelled region of the phase diagram has neither magnetic nor charge order.

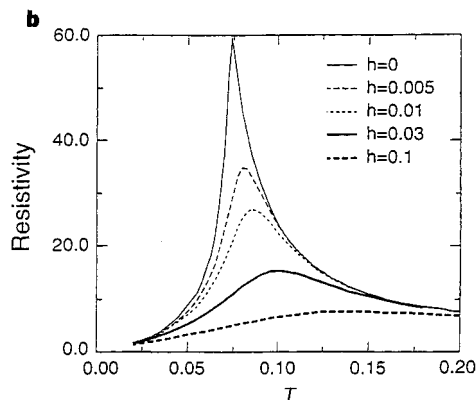
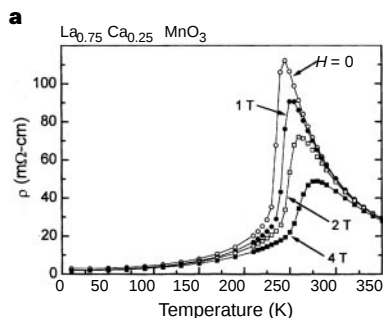


Figure 3 Measured and calculated resistivity of $\text{La}_{0.75}\text{Ca}_{0.25}\text{MnO}_3$. **a**, Measured resistivity at different magnetic fields: from ref. 7. **b**, Calculated resistivity (arbitrary units) versus temperature T in different magnetic fields h , from ref. 16. $h = 0.01$

corresponds to a field of 6 T. $T = 0.1$ corresponds to a temperature of $1/40$ of the bandwidth used in the calculation.

having its spin parallel to S_i^z so ferromagnetic order maximizes the hopping and antiferromagnetic order minimizes it. This connection between magnetic correlations and transport is called "double-exchange"⁸ and has been extensively studied^{9,10}.

Double-exchange provides a mechanism for the resistivity to change on passing through the paramagnetic–ferromagnetic transition: in the high-temperature phase the spins are disordered and scatter electrons, while in the low-temperature phase the spins are ordered, and do not scatter electrons. The difficulty is that even maximal spin disorder does not scatter electrons very much⁹. The spin scattering increases the resistivity somewhat, but (as has been shown most convincingly in recent numerical work¹¹) it localizes only a negligible fraction of the carriers in the band. Additional physical processes must therefore be invoked to explain the observed insulating behaviour. I will argue here that the crucial additional process is electron–lattice coupling.

Electron–lattice coupling

There are two sorts of electron–lattice coupling. One, called 'tolerance factor', involves the effect of the static crystal structure on electron hopping. In $\text{Re}_{1-x}\text{A}_x\text{MnO}_3$, different choices for Re and A have different atomic sizes and therefore produce different internal stresses acting on the Mn–O–Mn bonds. If the bond is compressed, it reacts by buckling, and the effective Mn–Mn hopping is a very sensitive function of this buckling, decreasing rapidly as the buckling increases. The change in average hopping amplitude as a function of average Re/A size has been studied in detail¹².

The other electron–lattice coupling is the conventional dynamical electron–phonon coupling, which links instantaneous deviations of atoms from their ideal crystallographic positions to the instantaneous deviations of electron configuration from the average values. There are three important modes: a 'breathing' distortion of the O_6 octahedron around a given Mn which couples to changes in the e_g charge density, and two linearly independent even-parity uniaxial volume-preserving distortions (Jahn–Teller modes) which couple to preferential occupancy of one e_g orbital over the other. One such distortion, indicated by arrows on Fig. 1a, involves a stretching of the Mn–O bonds along z and compression in the x – y plane and is favoured by occupancy of the ' $3z^2 - r^2 > e_g$ ' orbital. The corresponding change in the energy levels is shown in Fig. 1b.

A strong electron–phonon coupling may localize carriers, because the presence of an electron in a given Mn orbital causes a local lattice distortion which produces a potential minimum: this

minimum tends to trap the electron in that orbital. If the coupling is strong enough, these tendencies lead to the formation of a 'self-trapped' state called a polaron. However, according to the Pauli principle, two electrons can only occupy the same orbital if they have opposite spin; in the manganites this is prevented by the strong Hund's coupling, so 'bipolarons' do not occur. I note also that the usual polaron problem involves a single electron coupled to a deformable medium. The novelty of the manganites is the occurrence of self-trapping at a high density of electrons.

Self-trapping competes with the delocalizing tendency of electron hybridization. The competition is parametrized by a dimensionless quantity λ which is the ratio of the energy E_{latt} gained from the electron–phonon coupling in the absence of hybridization to the 'bare' electron kinetic energy t_{eff} ; thus $\lambda \approx E_{\text{latt}}/t_{\text{eff}}$. In the manganites, the nature of the double-exchange and the tolerance factor means that t_{eff} and therefore λ may be changed over a wide range by varying magnetic field and temperature, which change spin correlations, and also by changing Re and A, which change the tolerance factor and carrier concentration. Several recent theoretical papers^{13–16} have shown that the change in λ may be large enough to affect dramatically the behaviour of the materials, causing, for example, the high-temperature 'insulating' behaviour of $\text{La}_{0.75}\text{Ca}_{0.25}\text{MnO}_3$. The argument is that in the high-temperature state t_{eff} is sufficiently small that λ is so large that the electron–phonon interaction localizes the electrons; as T is decreased through $T_c(x)$, the growing ferromagnetic order increases t_{eff} and thus decreases λ sufficiently that metallic behaviour occurs. One calculation¹⁵ (based on these ideas) of the field and temperature dependence of the resistivity is shown in Fig. 3b; the qualitative agreement with the data suggests the calculation contains the correct physics.

Evidence for strong electron–lattice coupling

Jahn–Teller effects are known to be strong in at least some members of the family. For example, LaMnO_3 is an insulator with one e_g electron per Mn. In this compound the Jahn–Teller coupling causes an 'antiferrodistortive' deviation from the basic perovskite structure, in which some Mn–O bond lengths decrease and others increase, in a ' $(\pi, \pi, 0)$ ' pattern which alternates from Mn-site to Mn-site throughout the crystal¹⁷. Because the changes in bond length are long-range ordered and coherent throughout the crystal, they may be determined via conventional Bragg diffraction experiments: and have been found to be large: $\sim 0.2 \text{ \AA}$ or 10% of the mean Mn–O distance, implying a very strong Jahn–Teller coupling in LaMnO_3 . As x is increased from 0, the structural transition temperature and the size of the coherent, long-range structural distortion both decrease rapidly¹⁷. However, the theoretical models discussed above^{15,16} predicted that strong local distortions would persist over a wide range of doping, in the $T > T_c$ 'insulating' phase.

Recent advances in experimental technique, especially improvements in the measurement of extended X-ray absorption fine structure (EXAFS) and neutron pair distribution functions (PDF), have produced striking evidence for the predicted existence of local lattice distortions. Figure 4 shows the results of EXAFS measurements by Booth *et al.*¹⁸ of the root-mean-square variation in Mn–O bond length for different members of the $\text{La}_{1-x}\text{Ca}_x\text{MnO}_3$ series. The open symbols at the top of the figure are data for LaMnO_3 where, as discussed above, strong long-range distortions occur. The open symbols at the bottom of the figure pertain to CaMnO_3 , where no distortion at all occurs because the e_g band is empty. At intermediate dopings and high temperatures the variation in the bond-length distribution is of the order of the value found in the long-range ordered material; this indicates that large-amplitude local distortions, of a size comparable to that in the undoped material, exist even though no coherent part is visible in conventional Bragg scattering. As temperature is lowered through T_c , the distortion drops to a value typical of the undistorted

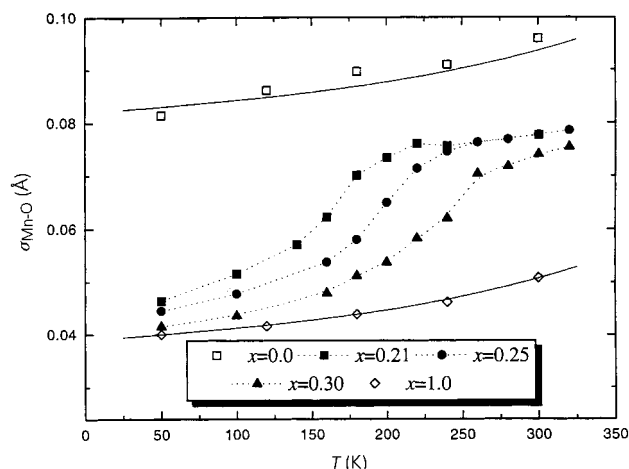


Figure 4 Temperature dependence of variance of Mn–O bond length for $\text{La}_{1-x}\text{Ca}_x\text{MnO}_3$ at various concentrations x , as determined by Booth *et al.*¹⁸. In ref. 18 these data are plotted as σ^2 versus T .

material. Other authors have obtained qualitatively similar results^{19,20}, although the details of the data and interpretations differ.

Additional experimental results consistent with the importance of an electron–lattice coupling include optical conductivity²¹, the large oxygen-isotope effect on the ferromagnetic T_c (ref. 22), the strong pressure dependence of physical properties^{23–25}, and the ubiquity at $x \geq 0.5$ of ‘charge-ordered’ phases accompanied by large changes in Mn–O bond lengths²⁶.

Challenges and opportunities

The manganites are a richly varied and fascinating system. Although a suggestive qualitative agreement between theory and experiment exists, much more needs to be done. Variability of data is a significant issue. For example, the optical data of ref. 21 are qualitatively similar to those reported by other workers²⁷, but differ substantially in detail. Also, as discussed in refs 15, 16 for instance, the theoretical calculations are oversimplified and a proper treatment including doping dependence, intersite correlations and the Coulomb interaction is urgently needed.

The materials also offer the intriguing possibility of establishing an explicit connection between atomic-scale ‘correlated electron’ physics and such materials-science questions as the effect of strain on physical properties. The hydrostatic pressure dependence of the ferromagnetic T_c has been measured²⁵, but if Jahn–Teller effects are important then the effects of volume-preserving uniaxial strain will be much larger. Preliminary work²⁸ supports this idea, and suggests that a quantitative understanding of strain effects may be possible.

Issues arising in the context of the manganites may be more generally important. One particularly significant theme is the relation between electron correlations, local structure, long-range order and physical properties. In the manganites there is clearly an interesting interplay between these phenomena, and connections have also been suggested for many other materials, including the high- T_c superconductors. Sophisticated techniques for examining local correlations are becoming available, and the manganites seem an ideal system for developing and validating them. This, and the intrinsic interest of the materials, should stimulate research for a long time to come. □

A. J. Millis is in the Department of Physics and Astronomy, The Johns Hopkins University, 3400 North Charles Street, Baltimore, Maryland 21218, USA

1. Wollan, E. D. & Koehler, W. C. Neutron diffraction study of the magnetic properties of the series of perovskite-type compounds $[(1-x)\text{La}, x\text{Ca}]\text{MnO}_3$. *Phys. Rev.* **100**, 545–563 (1955).

2. von Helmholtz, R. M., Wecker, J., Holzapfel, B., Schultz, L. & Samwer, K. Giant negative magnetoresistance in perovskite-like $\text{La}_{2/3}\text{Ba}_{1/3}\text{MnO}_x$ ferromagnetic films. *Phys. Rev. Lett.* **71**, 2331–2333 (1993).
3. Jin, S. *et al.* Thousandfold change in resistivity in magnetoresistive La–Ca–Mn–O films. *Science* **264**, 413–415 (1994).
4. Wu, M. W. & Conwell, E. M. Transport in alpha-sexithiophene films. *Chem. Phys. Lett.* **266**, 363–367 (1997).
5. Teslic, S., Egami, T. & Viehland, D. Local atomic structure of PZT and PLZT studied by pulsed neutron scattering. *J. Phys. Chem. Solids* **57**, 1537–1543 (1997).
6. Terminello, L. J., Mini, S. M., Ade, H. & Perry, D. L. (eds) *Applications of Synchrotron Radiation Techniques to Materials Science III* (Mater. Res. Soc., Pittsburgh, PA, 1997).
7. Schiffer, P., Ramirez, A. P., Bao, W. & Cheong, S.-W. Low temperature magnetoresistance and the magnetic phase diagram of $\text{La}_{1-x}\text{Ca}_x\text{MnO}_3$. *Phys. Rev. Lett.* **75**, 3336–3339 (1995).
8. Anderson, P. W. & Hasegawa, H. Considerations on double-exchange. *Phys. Rev.* **100**, 675–681 (1955).
9. Kubo, K. & Ohata, A. A quantum theory of double-exchange. *J. Phys. Soc. Jpn* **33**, 21–32 (1972).
10. Nagaev, E. L. Lanthanum manganites and other giant-magnetoresistance magnetic conductors. *Usp. Fiz. Nauk* **166**, 833–858 (1996); *Physics-Uspekhi* **39**, 781 (1996).
11. Li, Q., Zang, J., Bishop, A. R. & Soukoulis, C. M. Charge localization in disordered colossal-magnetoresistance manganites. *Phys. Rev. B* **56**, 4541–4544 (1997).
12. Hwang, H.-Y., Cheong, S.-W., Raedelli, P. G., Marezio, M. & Batlogg, B. Lattice effects on the magnetoresistance in doped LaMnO_3 . *Phys. Rev. Lett.* **75**, 914–917 (1995).
13. Roder, H., Zang, J. & Bishop, A. R. Lattice effects in the colossal-magnetoresistance manganites. *Phys. Rev. Lett.* **76**, 1356–1359 (1996).
14. Zang, J., Bishop, A. R. & Roder, H. Double degeneracy and Jahn–Teller effects in colossal-magnetoresistance perovskites. *Phys. Rev. B* **53**, 8840–8843 (1996).
15. Millis, A. J., Shraiman, B. I. & Mueller, R. Dynamic Jahn–Teller effect and colossal magnetoresistance in $\text{La}_{1-x}\text{Sr}_x\text{MnO}_3$. *Phys. Rev. Lett.* **77**, 175–178 (1996).
16. Millis, A. J., Mueller, R. & Shraiman, B. I. Fermi-liquid-to-polaron crossover. II. Double exchange and the physics of colossal magnetoresistance. *Phys. Rev. B* **54**, 5405–5417 (1996).
17. Ellemans, J. B. A., vanLaar, B., vanderVeer, K. J. R. & Loopstra, B. O. The crystallographic and magnetic structures of $\text{La}_{1-x}\text{Ba}_x\text{Mn}_{1-y}\text{Me}_y\text{O}_3$. *J. Solid State Chem.* **3**, 238–242 (1971).
18. Booth, C. H. *et al.* Direct relationship between magnetism and MnO_6 distortions in $\text{La}_{1-x}\text{Ca}_x\text{MnO}_3$. *Phys. Rev. Lett.* **80**, 853 (1998).
19. Billinge, S. J. L., DiFrancesco, R. G., Kwei, G. H., Neumeier, J. J. & Thompson, J. D. Direct observation of lattice polaron formation in the local structure of $\text{La}_{1-x}\text{Ca}_x\text{MnO}_3$. *Phys. Rev. Lett.* **77**, 715–718 (1996).
20. Louca, D., Egami, T., Brousha, E. L., Roder, H. & Bishop, A. R. Local Jahn–Teller distortion in $\text{La}_{1-x}\text{Sr}_x\text{MnO}_3$ observed by pulsed neutron diffraction. *Phys. Rev. B* **56**, 8475–8478 (1997).
21. Kaplan, S. G. *et al.* Optical evidence for the dynamic Jahn–Teller effect in $\text{Nd}_{0.5}\text{Sr}_{0.5}\text{MnO}_3$. *Phys. Rev. Lett.* **77**, 2081–2084 (1996).
22. Zhao, G.-M., Conder, K., Keller, H. & Muller, K. A. Giant oxygen isotope shift in the magnetoresistive perovskite $\text{La}_{1-x}\text{Ca}_x\text{MnO}_{3+y}$. *Nature* **381**, 676–678 (1996).
23. Khazeni, K. *et al.* Effect of pressure on the magnetoresistance of single crystal $\text{Nd}_{0.5}\text{Sr}_{0.5}\text{Pb}_{0.14}\text{MnO}_{3-\delta}$. *Phys. Rev. Lett.* **76**, 295–298 (1996).
24. Moritomo, Y., Asamitsu, A. & Tokura, Y. Pressure effect on the double-exchange ferromagnet $\text{La}_{1-x}\text{Sr}_x\text{MnO}_3$. *Phys. Rev. B* **51**, 16491–16494 (1996).
25. Neumeier, J. J., Hundley, M. E., Thompson, J. D. & Heffner, R. H. Substantial pressure effects on the electrical resistivity and ferromagnetic transition temperature of $\text{La}_{1-x}\text{Ca}_x\text{MnO}_3$. *Phys. Rev. B* **52**, 7006–7009 (1995).
26. Radaelli, P. G., Cox, D. E., Marezio, M. & Cheong, S.-W. Charge, orbital, and magnetic ordering in $\text{La}_{0.5}\text{Ca}_{0.5}\text{MnO}_3$. *Phys. Rev. B* **55**, 3015–3023 (1997).
27. Okimoto, Y. *et al.* Anomalous variation of optical spectra with spin polarization in double-exchange ferromagnet: $\text{La}_{1-x}\text{Sr}_x\text{MnO}_3$. *Phys. Rev. Lett.* **75**, 109–112 (1995).
28. Millis, A. J., Darling, T. & Migliori, A. Quantifying the effects of strain in ‘colossal’ magnetoresistance manganites. *J. Appl. Phys.* **83**, 1588–1591 (1998).

Acknowledgements. I thank S.-W. Cheong for Fig. 3 and K. Ahn for assistance with Fig. 1. This work was supported by the US NSF.

Correspondence should be addressed to the author (e-mail: millis@pha.jhu.edu).

Observation of Feshbach resonances in a Bose–Einstein condensate

S. Inouye*, M. R. Andrews*†, J. Stenger*, H.-J. Miesner*, D. M. Stamper-Kurn* & W. Ketterle*

* Department of Physics and Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

It has long been predicted that the scattering of ultracold atoms can be altered significantly through a so-called ‘Feshbach resonance’. Two such resonances have now been observed in optically trapped Bose–Einstein condensates of sodium atoms by varying an external magnetic field. They gave rise to enhanced inelastic processes and a dispersive variation of the scattering length by a factor of over ten. These resonances open new possibilities for the study and manipulation of Bose–Einstein condensates.

Bose–Einstein condensates of atomic gases offer new opportunities for studying quantum-degenerate fluids^{1–5}. All the essential properties of Bose condensed systems—the formation and shape of the condensate, the nature of its collective excitations and statistical fluctuations, and the formation and dynamics of solitons and vortices—are determined by the strength of the atomic interactions. In contrast to the situation for superfluid helium, these interactions are weak, allowing the phenomena to be theoretically described from ‘first principles’. Furthermore, in atomic gases the interactions can be altered, for instance by employing different species, changing the atomic density, or, as in the present work, merely by varying a magnetic field.

At low temperatures, the interaction energy in a cloud of atoms is proportional to the density and a single atomic parameter, the scattering length a which depends on the quantum-mechanical phase shift in an elastic collision. It has been predicted that the scattering length can be modified by applying external magnetic^{6–10}, optical^{11,12} or radio-frequency¹³ (r.f.) fields. Those modifications are only pronounced in a so-called ‘Feshbach resonance’¹⁴, when a quasibound molecular state has nearly zero energy and couples resonantly to the free state of the colliding atoms. In a time-dependent picture, the two atoms are transferred to the quasibound state, ‘stick’ together and then return to an unbound state. Such a resonance strongly affects the scattering length (elastic channel), but also affects inelastic processes such as dipolar relaxation^{6,7} and three-body recombination. Feshbach resonances have so far been studied at much higher energies¹⁵ by varying the collision energy, but here we show that they can be ‘tuned’ to zero energy to be resonant for ultracold atoms. The different magnetic moments of the free and quasibound states allowed us to tune these resonances with magnetic fields, and as a result, minute changes in the magnetic field strongly affected the properties of a macroscopic system.

Above and below a Feshbach resonance, the scattering length a covers the full continuum of positive and negative values. This should allow the realization of condensates over a wide range of interaction strengths. By setting $a \approx 0$, one can create a condensate with essentially non-interacting atoms, and by setting $a < 0$ one can make the system unstable and observe its collapse. Rapid tuning of an external magnetic field around a Feshbach resonance will lead to sudden changes of the scattering length. This opens the way to studies of new dynamical effects such as novel forms of collective oscillations or the sudden collapse of a large condensate when the scattering length is switched from positive to negative¹⁶.

Theoretical predictions

Calculations for Feshbach resonances in external magnetic fields have been reported for the lower hyperfine states of the atoms Li (ref. 8), K (ref. 10), Na (ref. 8), Rb (ref. 9) and Cs (refs 6, 7). They are typically spaced by several hundred gauss, and for Li and Na occur outside the range where states in the lower hyperfine manifold are weak-field-seeking and can be magnetically trapped. Recent experimental efforts to observe Feshbach resonances have concentrated on ⁸⁷Rb (ref. 17) and on ⁸⁵Rb (ref. 18 and C. E. Wieman, personal communication) where Feshbach resonances have been predicted at relatively low magnetic fields⁹. However, our recently demonstrated all-optical confinement of a Bose condensate¹⁹ opened the possibility of observing Feshbach resonances for strong-field-seeking states which cannot be trapped in a d.c. magnetic trap. The optical trapping potential is unaffected by magnetic fields and is independent of the hyperfine ground state. We report here the observation of two Feshbach resonances of sodium in a strong-field-seeking state.

Several Feshbach resonances in sodium are caused by quasibound hyperfine states of the second highest vibrational level, $\nu = 14$, of the triplet potential of the sodium dimer. The lowest magnetic field value B_0 for a strong Feshbach resonance in sodium was predicted to lie in the range $760 < B_0 < 925$ G (B. J. Verhaar and F. A. van Abeelen, personal communication). It occurs in collisions between atoms in the lowest hyperfine state $|m_s = -1/2, m_l = +3/2\rangle$, which correlates with the $|F = 1, m_F = +1\rangle$ state at low fields (S , I and F are the usual quantum numbers for the electronic, nuclear and total spin, respectively). This Feshbach resonance is due to a quasibound molecular state $|S = 1, m_s = +1, I = 1, m_l = +1\rangle$. A much weaker resonance due to a $|S = 1, m_s = +1, I = 3, m_l = +1\rangle$ state (which is almost degenerate with the other quasibound state) was predicted to occur 50 to 75 G below.

Near a Feshbach resonance, the scattering length a should vary dispersively as a function of magnetic field B (ref. 8):

$$a = \bar{a} \left(1 - \frac{\Delta}{B - B_0} \right) \quad (1)$$

where Δ parametrizes the width of the resonance at $B = B_0$, and $\bar{a}$ is the scattering length outside the resonance. For sodium, $\bar{a}$ was found spectroscopically to be 2.75 nm at zero field, and increases to the triplet scattering length of 4.5 nm (ref. 20) at high magnetic fields. The widths Δ for the strong and weak resonance were predicted to be 1 G and 0.01 G, respectively (B. J. Verhaar and F. A. van Abeelen, personal communication).

† Present address: Bell Laboratories, Lucent Technologies, Murray Hill, New Jersey 07974, USA.

Experimental set-up

Bose–Einstein condensates in the $|F = 1, m_F = -1\rangle$ state were produced as in our previous work by laser cooling, followed by evaporative cooling in a magnetic trap²¹. The condensates were transferred into an optical dipole trap formed at the focus of an infrared laser beam¹⁹. Atoms were then spin-flipped with nearly 100% efficiency to the $|F = 1, m_F = +1\rangle$ state with an adiabatic r.f. sweep while applying a 1 G bias field. Without large modifications of our magnetic trapping coils, we could provide bias fields of up to $\sim 1,200$ G, but only by using coils producing axial curvature²¹, which for high-field-seeking states generated a repulsive axial potential. At the highest magnetic fields, this repulsion was stronger than the axial confinement provided by the optical trap. To prevent the atoms from escaping, two ‘end-caps’ of far-off-resonant blue-detuned laser light were placed at the ends of the condensate, creating a repulsive potential, and confining the atoms axially (Fig. 1a). For this, green light at 514 nm from an argon-ion laser was focused into two sheets about 200 μm apart. The focus of the optical trap was placed near the minimum of the bias field in order to minimize the effect of the destabilizing magnetic field curvature. The axial trapping potential at high fields was approximately ‘W’-shaped (Fig. 1b), and had a minimum near one of the end-caps as observed by phase-contrast imaging²² (Fig. 1c, d).

The calibration factor between the current (up to ~ 400 A) in the coils and the magnetic bias field was determined with an accuracy of 2% by inducing r.f. transitions within the $|F = 1\rangle$ ground-state hyperfine manifold at about 40 G. Additionally, an optical resonance was found around 1,000 G, where the Zeeman shift equalled the probe light detuning of about 1.7 GHz and led to a sign-reversal of the phase-contrast signal. These two calibrations agreed within their uncertainties.

The condensate was observed in the trap directly using phase-contrast imaging²² or by using time-of-flight absorption imaging^{1,2,21}. In the latter case, the optical trap was suddenly switched off, and the magnetic bias field was shut off 1–2 ms later to ensure that the high-field value of the scattering length was responsible for the acceleration of the atoms. After ballistic expansion of the condensate (either 12 or 20 ms), the atoms were optically pumped into the $|F = 2\rangle$ ground state and probed using resonant light driving the cycling transition. The disk-like expansion of the

cloud and the radial parabolic density profile were clear evidence for the presence of a Bose condensed cloud.

Locating the resonances

When the magnetic field is swept across a Feshbach resonance one would expect to lose a condensate due to an enhanced rate of inelastic collisions (caused either by the collapse in the region of negative scattering length or by an enhanced rate coefficient for inelastic collisions). This allowed us to implement a simple procedure to locate the resonances: we first extended the field ramp until the atoms were lost and then used successively narrower field intervals to localize the loss. This procedure converged much faster than a point-by-point search. As we could take many non-destructive phase-contrast images during the magnetic field ramp, the sharp onset of trap loss at the resonance was easily monitored (Fig. 1c, d).

The most robust performance was obtained by operating the optical dipole trap at 10 mW laser power focused to a beam waist of 6 μm , resulting in tight confinement of the condensate and therefore rather short lifetimes owing to three-body recombination¹⁹. This required that the magnetic field be ramped up in two stages: a fast ramp at a rate of ~ 100 G ms^{-1} to a value slightly below that expected for a Feshbach resonance, followed by a slow ramp at a rate between 0.05 and 0.3 G ms^{-1} to allow for detailed observation. Near 907 G, we observed a dramatic loss of atoms, as shown in Figs 1c and 2a. This field value was reproducible to better than 0.5 G and had a calibration uncertainty of ± 20 G.

To distinguish between an actual resonance and a threshold for trap loss, we also approached the resonance from above. Fields above the Feshbach resonance were reached by ramping at a fast rate of 200 G ms^{-1} , thus minimizing the time spent near the resonance and the accompanying losses. The number of atoms above the resonance was typically three times smaller than below. Approaching the resonance from above, a similarly sharp loss phenomenon was observed about 1 G higher in field than from below (Fig. 2a), which roughly agrees with the predicted width of the resonance. A second resonance was observed 54 ± 1 G below the first one, with the observed onset of trap loss at least a factor of ten sharper than for the first. As the upper resonance was only reached by passing through the lower one, some losses of atoms were unavoidable; for example, when the lower resonance was crossed at 2 G ms^{-1} ,

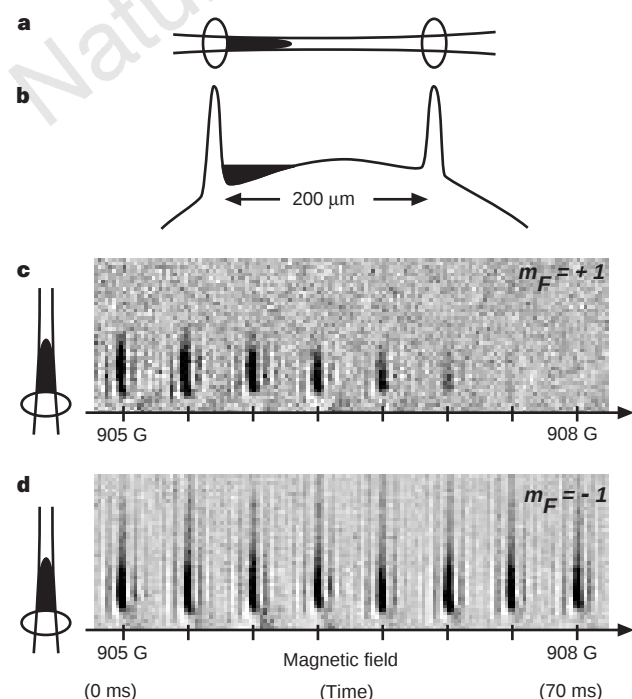


Figure 1 Observation of the Feshbach resonance at 907 G using phase-contrast imaging in an optical trap. A rapid sequence (100 Hz) of non-destructive, *in situ* phase-contrast images of a trapped cloud (which appears black) is shown. As the magnetic field was increased, the cloud suddenly disappeared for atoms in the $|m_F = +1\rangle$ state (see images in **c**), whereas nothing happened for a cloud in the $|m_F = -1\rangle$ state (images in **d**). The height of the images is 140 μm . A diagram of the optical trap is shown in **a**. It consisted of one red-detuned laser beam providing radial confinement, and two blue-detuned laser beams acting as end-caps (shown as ovals). The minimum of the magnetic field was slightly offset from the centre of the optical trap. As a result, the condensate (shaded area) was pushed by the magnetic field curvature towards one of the end-caps. The axial profile of the total potential is shown in **b**.

about 80% of the atoms were lost. This, coupled with the stability and finite programming speed of the power supplies, limited the ramp rates to those given above.

The observation of twin resonances separated by 54 ± 1 G, with the weaker one at lower field, exactly matches the theoretically predicted pattern and thus strongly confirms our interpretation. No resonance phenomena were observed in the $|m_F = -1\rangle$ state at any field up to 1,000 G, in agreement with theory which predicted resonances for this state only at much higher fields.

Changing the scattering length

The trap loss measurements easily located the Feshbach resonances. To measure the variation of the scattering length around these resonances, we determined the interaction energy of a trapped condensate. This was done by suddenly switching off the trap, allowing the stored interaction energy to be converted into the kinetic energy of a freely expanding condensate and measuring it by time-of-flight absorption imaging^{1,2,21}. The interaction energy is proportional to the scattering length and the average density of the condensate $\langle n \rangle$:

$$E_I/N = \frac{2\pi\hbar^2}{m} a \langle n \rangle \quad (2)$$

where N is the number of condensed atoms of mass m . For a large condensate the kinetic energy in the trap is negligible (Thomas–Fermi limit), and E_I is equal to the kinetic energy E_K of the freely expanding condensate $E_K/N = mv_{\text{rms}}^2/2$, where v_{rms} is the root-mean-square velocity of the atoms. For a three-dimensional har-

monic oscillator potential one finds $\langle n \rangle \propto N(Na)^{-3/5}$ (ref. 23) (We note that, for a general power-law potential $\Sigma_i c_i x_i^{\beta_i}$, one obtains $\langle n \rangle \propto N(Na)^{k-1}$, where $k = 1/(1 + \Sigma_i 1/\beta_i)$). Thus, the value of the scattering length scales as:

$$a \propto \frac{v_{\text{rms}}^5}{N} \quad (3)$$

Both v_{rms} and N can be directly evaluated from absorption images of freely expanding condensates. For a cigar-shaped condensate the free expansion is predominantly radial, and so the contribution of the axial dimension to v_{rms} could be neglected. The quantity v_{rms}^2/N (equation (3)), normalized to unity outside the resonance, should be identical to $a/\bar{a}$ (equation (1)). This quantity was measured around the resonance at 907 G and is shown in Fig. 2b together with the theoretical prediction of a resonance with width $\Delta = 1$ G. The data clearly displays the predicted dispersive shape and shows evidence for a variation in the scattering length by more than a factor of ten.

We now discuss the assumptions for equation (3) and show that it is approximately valid for our conditions. (1) We assumed that the condensate remains in equilibrium during the magnetic field ramp. This is the case if the adiabatic condition $\dot{a}/a \ll \omega_i$ holds for the temporal change of the scattering length¹⁶, and a similar condition for the loss of atoms (the ω_i are the trapping frequencies). For the condensate's fast radial dynamics ($\omega_r \approx 2\pi \times 1.5$ kHz) this condition is fulfilled, whereas for the slower axial motion ($\omega_z \approx 2\pi \times 0.1$ kHz) it breaks down close to or within the resonance. In this case the density would approach the two-dimensional scaling $N(Na)^{-1/2}$, but the values for $a/\bar{a}$ (Fig. 2b) would differ by at most 50%. (2) The second assumption was a three-dimensional harmonic trap. If the axial potential has linear contributions, the density scales instead like $N(Na)^{-2/3}$ resulting in at most a 50% change for $a/\bar{a}$. (3) We assumed that contributions of collective excitations to the released energy were small. Axial striations were observed in free expansion for both $|m_F = +1\rangle$ and $|m_F = -1\rangle$ atoms (probably created by the changing potential during the fast magnetic field ramp). However, the small scatter of points outside the resonance in Fig. 2b, which do not show any evidence of oscillations, suggests that the contribution of excitations to the released energy is negligible. (4) We assumed a sudden switch-off of the trap and ballistic expansion. The inhomogeneous bias field during the first 1–2 ms of free expansion accelerated the axial expansion, but had a negligible effect on the expansion of the condensate in the radial direction, which was evaluated for Fig. 2b.

None of the corrections (1)–(4) discussed above affect our conclusion that the scattering length varies dispersively near a Feshbach resonance. More accurate experiments should be done with a homogeneous bias field. In addition, an optical trap with larger volume and lower density would preclude the need to ramp the field quickly because three-body recombination would be reduced.

The trap losses observed around the Feshbach resonances merit further study as they might impose practical limits on the possibilities for varying the scattering length. An increase of the dipolar relaxation rate near Feshbach resonances has been predicted^{6,7}, but for atoms in the lowest hyperfine state no such inelastic binary collisions are possible. Therefore, the observed trap loss is probably due to three-body collisions. In this case the loss rate is characterized by the coefficient K_3 , defined as $\dot{N}/N = -K_3 \langle n^2 \rangle$. So far, there is no theoretical work on K_3 near a Feshbach resonance. An analysis based on Fig. 2 shows that K_3 increased on both sides of the resonance, because the loss rate increased while the density decreased or stayed constant. In any case, the fact that we observed Feshbach resonances at high atomic densities ($\sim 10^{15} \text{ cm}^{-3}$) strongly enhanced this loss process, which can be avoided with a condensate at lower density in a modified optical trap. Control of the bias field with a precision better than $\sim 10^{-4}$ will be necessary to achieve negative or extremely large values of the scattering length in a stable way.

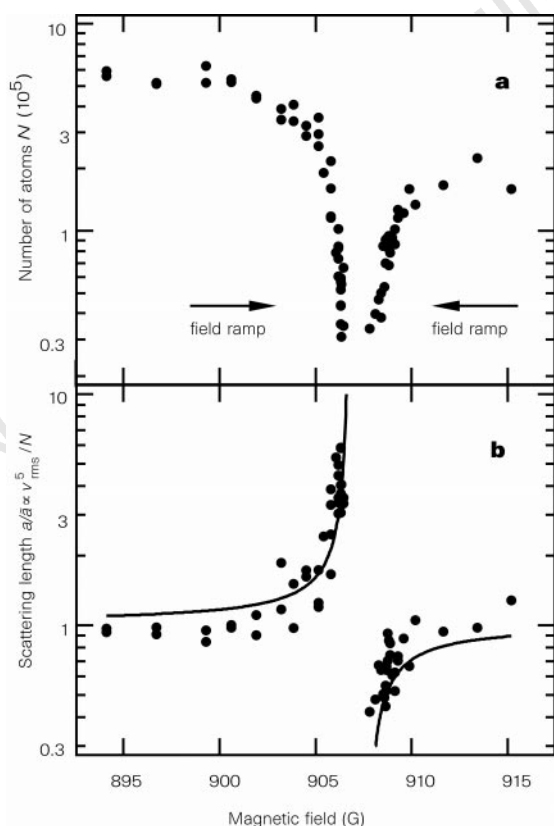


Figure 2 Observation of the Feshbach resonance at 907 G using time-of-flight absorption imaging. **a**, Number of atoms in the condensate versus magnetic field. Field values above the resonance were reached by quickly crossing the resonance from below and then slowly approaching from above. **b**, The normalized scattering length $a/\bar{a} \propto v_{\text{rms}}^5/N$ calculated from the released energy, together with the predicted shape (equation (1), solid line). The values of the magnetic field in the upper scan relative to the lower one have an uncertainty of <0.5 G.

A tunable condensate

We have observed two Feshbach resonances for Bose–Einstein condensates of sodium through the abrupt loss of atoms, and obtained strong evidence for a dispersive variation of the scattering length by a factor of more than ten. ‘Tuning’ of the scattering length should become an important tool for ‘designing’ atomic quantum gases with novel properties; for example, to create ideal Bose–Einstein condensates with nearly zero scattering length, and to obtain a detailed picture of the collapse of a condensate with negative scattering length, which is so far not fully understood. Tuning the scattering length can also be used to vary interactions between different species²⁴ and thus control the phase diagram of multi-component condensates, possibly switching from interpenetrating superfluids to phase separation²⁵. Feshbach resonances may also be important in atom optics, for modifying the atomic interactions in an atom laser, or more generally, for controlling nonlinear coefficients in atom optics with coherent beams of atoms. □

Received 18 February; accepted 19 February 1998.

1. Anderson, M. H., Ensher, J. R., Matthews, M. R., Wieman, C. E. & Cornell, E. A. Observation of Bose–Einstein condensation in a dilute atomic vapor. *Science* **269**, 198–201 (1995).
2. Davis, K. B. *et al.* Bose–Einstein condensation in a gas of sodium atoms. *Phys. Rev. Lett.* **75**, 3969–3973 (1995).
3. Bradley, C. C., Sackett, C. A. & Hulet, R. G. Bose–Einstein condensation of lithium: Observation of limited condensate number. *Phys. Rev. Lett.* **78**, 985–989 (1997).
4. Georgia Southern University BEC home page, <http://amo.phy.gasou.edu/bec.html>.
5. Bradley, C. C., Sackett, C. A., Tollett, J. J. & Hulet, R. G. Evidence of Bose–Einstein condensation in an atomic gas with attractive interactions. *Phys. Rev. Lett.* **75**, 1687–1690 (1995).
6. Tiesinga, E., Moerdijk, A. J., Verhaar, B. J. & Stoof, H. T. C. Conditions for Bose–Einstein condensation in magnetically trapped atomic cesium. *Phys. Rev. A* **46**, R1167–R1170 (1992).
7. Tiesinga, E., Verhaar, B. J. & Stoof, H. T. C. Threshold and resonance phenomena in ultracold atomic state collisions. *Phys. Rev. A* **47**, 4114–4122 (1993).
8. Moerdijk, A. J., Verhaar, B. J. & Axelsson, A. Resonances in ultracold collisions of ⁶Li, ⁷Li and ²³Na. *Phys. Rev. A* **51**, 4852–4861 (1995).

9. Vogels, J. M. *et al.* Prediction of Feshbach resonances in collisions of ultracold rubidium atoms. *Phys. Rev. A* **56**, R1067–R1070 (1997).
10. Boesten, H. M. J. M., Vogels, J. M., Tempelaars, J. G. C. & Verhaar, B. J. Properties of cold collisions of ³⁹K atoms and of ⁴¹K atoms in relation to Bose–Einstein condensation. *Phys. Rev. A* **54**, R3726–R3729 (1996).
11. Fedichev, P. O., Kagan, Yu., Shlyapnikov, G. V. & Walraven, J. T. M. Influence of nearly resonant light on the scattering length in low-temperature atomic gases. *Phys. Rev. Lett.* **77**, 2913–2916 (1996).
12. Bohn, J. L. & Julienne, P. S. Prospects for influencing the scattering lengths with far-off-resonant light. *Phys. Rev. A* **56**, 1486–1491 (1997).
13. Moerdijk, A. J., Verhaar, B. J. & Nagtegaal, T. M. Collisions of dressed ground-state atoms. *Phys. Rev. A* **53**, 4343–4351 (1996).
14. Feshbach, H. A unified theory of nuclear reactions. II. *Ann. Phys.* **19**, 287–313 (1962).
15. Bryant, H. C. *et al.* Observation of resonances near 11 eV in the photodetachment cross-section of the H⁺ ion. *Phys. Rev. Lett.* **38**, 228–230 (1977).
16. Kagan, Yu., Surkov, E. L. & Shlyapnikov, G. V. Evolution and global collapse of trapped Bose condensates under variations of the scattering length. *Phys. Rev. Lett.* **79**, 2604–2607 (1997).
17. Newbury, N. R., Myatt, C. J. & Wieman, C. E. s-wave elastic collisions between cold ground-state ⁸⁷Rb atom. *Phys. Rev. A* **51**, R2680–R2683 (1995).
18. Courteille, P. & Heinzen, D. Paper presented at SPIE Photonics West, 24–30 Jan., San Jose, California (1998).
19. Stemper-Kurn, D. M. *et al.* Optical confinement of a Bose–Einstein condensate. *Phys. Rev. Lett.* (in the press).
20. Tiesinga, E. *et al.* A spectroscopic determination of scattering lengths for sodium atom collisions. *J. Res. Natl Inst. Stand. Technol.* **101**, 505–520 (1996).
21. Mewes, M.-O. *et al.* Bose–Einstein condensation in a tightly confining d.c. magnetic trap. *Phys. Rev. Lett.* **77**, 416–419 (1996).
22. Andrews, M. R. *et al.* Propagation of sound in a Bose–Einstein condensate. *Phys. Rev. Lett.* **79**, 553–556 (1997).
23. Baym, G. & Pethick, C. J. Ground-state properties of magnetically trapped Bose-condensed rubidium gas. *Phys. Rev. Lett.* **76**, 6–9 (1996).
24. van Abeelen, F. A., Verhaar, B. J. & Moerdijk, A. J. Sympathetic cooling of ⁶Li atoms. *Phys. Rev. A* **55**, 4377–4381 (1997).
25. Ho, T.-L. & Shenoy, V. B. Binary mixtures of Bose condensates of alkali atoms. *Phys. Rev. Lett.* **77**, 3276–3279 (1996).

Acknowledgements. We thank J. M. Vogels for discussions, A. P. Chikkatur for experimental assistance, and B. J. Verhaar and F. A. van Abeelen for providing updated theoretical predictions. We also thank D. Kleppner, D. E. Pritchard and R. A. Rubenstein for a critical reading of the manuscript. This work was supported by the Office of Naval Research, NSF, Joint Services Electronics Program (ARO), and the David and Lucile Packard Foundation. J.S. acknowledges support from the Alexander von Humboldt-Foundation, and D.M.S.-K. was supported by the JSEP graduate fellowship program.

Correspondence and requests for materials should be addressed to W.K.

YOURS TO HAVE AND TO HOLD BUT NOT TO COPY

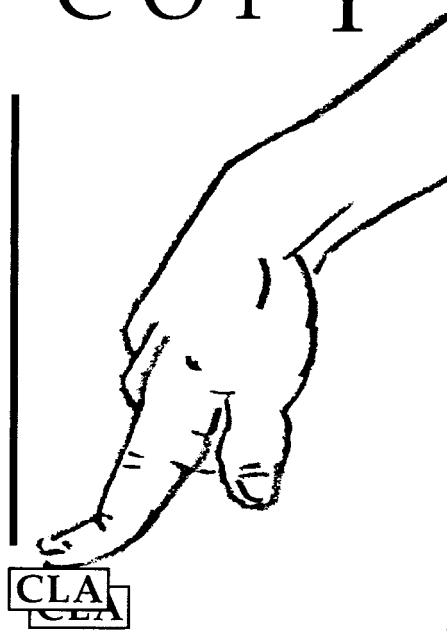
The publication you are reading is protected by copyright law. This means that the publisher could take you and your employer to court and claim heavy legal damages if you make unauthorised infringing photocopies from these pages.

Photocopying copyright material without permission is no different from stealing a magazine from a newsagent, only it doesn't seem like theft.

The Copyright Licensing Agency (CLA) is an organisation which issues licences to bring photocopying within the law. It has designed licensing services to cover all kinds of special needs in business, education, and government.

If you take photocopies from books, magazines and periodicals at work your employer should be licensed with CLA.

Make sure you are protected by a photocopying licence.



The Copyright Licensing Agency Limited
90 Tottenham Court Road, London W1P 0LP
Telephone: 0171 436 5931
Fax: 0171 436 3986

The Sun's shape and brightness

J. R. Kuhn^{*†}, R. I. Bush[‡], X. Scheick[§] & P. Scherrer[‡]

^{*} Michigan State University, East Lansing, Michigan 48823, USA

[†] National Solar Observatory, Sunspot, New Mexico 88349, USA

[‡] Stanford University, Stanford, California 94305, USA

[§] Jackson Community College, Jackson, Michigan 49201, USA

The origin of the 11- and 22-year solar cycles remains one of the more mysterious aspects of the Sun. These cycles are probably driven by convection in the solar interior, but the convection zone is difficult to probe. Small departures from sphericity in the effective surface temperature of the Sun can in principle be used in this regard. Such variations, which are observed as changes in the surface brightness with solar latitude, may be caused by differences between the vertical and horizontal turbulent convective flows¹ inside the Sun. Moreover, variations in the Sun's luminosity may be related to changes in conditions near the base of the convection zone² that result from the magnetic (sunspot) cycle³. Here we present satellite data that show that the Sun's shape and temperature vary with latitude in an unexpectedly complex way. Although the solar oblateness shows no evidence of varying with the solar cycle, we find a significant hexadecapole shape term which may vary. We also see a variation of about 1.5 K in the surface temperature with latitude. Based on these results, we suggest that sensitive observations of brightness variations be used as a record of the surface 'shadow' of cyclical changes in the solar interior.

Measurements of the shape of the Sun, and in particular the solar oblateness, have been debated for nearly a century⁴. Interest in the solar oblateness over the past 35 years was kindled by measurements⁵ that suggested that general relativity may be seriously 'incomplete'. Subsequent data^{6–10} did not confirm these results, although they have, on occasion, been interpreted as

evidence that the Sun's shape varies over solar cycle timescales¹¹. Such a variation, should it be confirmed, would require unexpectedly large solar interior rotation, or density, changes to occur over only a few years. This would have a profound effect on our understanding of the solar cycle.

The photometric and astrometric data described here were obtained with the Michelson doppler imager (MDI) experiment¹² on 19 March 1996 and 20 March 1997. To measure the limb shape and brightness, we decompose the solar limb position and brightness (obtained from full-disk MDI 1,024 × 1,024 pixel 'continuum' charge-coupled-device (CCD) images) into a linear combination of low order Legendre polynomials (P_l) of the form $\sum_{l=0,2,4} c_l P_l(\cos \theta')$ (where θ' is the solar colatitude). The $l = 2$ term corresponds to a quadrupolar distortion (the 'oblateness'). Both the quadrupole and hexadecapole ($l = 4$) shape terms are affected by the solar gravitational potential field and the solar rotation (and differential rotation).

It is critical that the spacecraft, and therefore the MDI optics, be rotated through 360 degrees by small angular increments, because the data must be analysed by a 'phase modulation' technique to extract the small solar limb shape and brightness signal from the much larger instrumental noise background. Although we achieve accuracy at the level of about half a kilometre in the limb shape, temporal variations of higher angular harmonics of the limb shape have been measured from MDI data with the even greater sensitivity of 10 microarcsec (corresponding to a few metres at the solar limb) over a two-month duration¹³.

The 1996 observations were obtained from a SOHO spacecraft commissioning roll, which included 90-degree roll angle increments with ~60-min dwell at each stop. The 1997 roll was performed in increments of 30 degrees with a 25-min dwell at each stop. The 1996 measurements were obtained without the active MDI image stabilization system which uses an active tip-tilt mirror because of uncertainties in the spacecraft pointing performance. During the 1997 roll, however, the MDI image stabilization system maintained the solar image position to better than 0.001 pixels. Our best data were obtained in 1997, in part because the 12-position roll provided a measurement of higher shape harmonics and the limb bright-

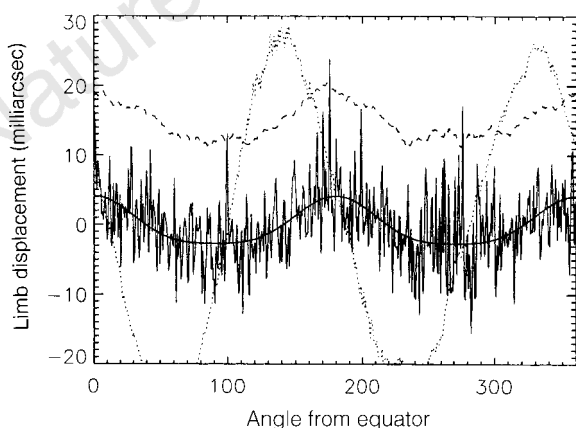


Figure 1 Limb displacement, $S(\theta)$, for each of the 512 angular bins around the Sun, as derived from 1997 measurements. From N evenly spaced roll angles distributed between 0 and 2π radians, it can be shown that $\sum_{k=0}^{N-1} \alpha_k(\theta_k) \approx N I_s(\theta)$. Here $\approx$ indicates that equality holds up to the $(N-1)$ th angular harmonic component of α , β , I , S and B . Once I_s and I_b are determined, the solar contributions S and B , can be computed. With only 4 roll angles (spaced 90 degrees apart) the solar shape and brightness harmonic components with $l = 0, \dots, 3$, which vary as $\cos(l\theta)$ or $\sin(l\theta)$ are determined. Harmonics with $l = 0$ and 1 are filtered from the data because of their sensitivity to the scale (focus) and displacement (jitter) of the image. Plotted data were obtained by averaging the derived β coefficients from all of the roll positions, shifted into the solar reference frame. The 4-milliarcsec r.m.s. high-frequency noise that is apparent in the figure may be due to residual photometric gain errors and not solar limb structure, as both the 1996 and 1997 data were obtained near solar minimum, at a time when there was very little solar magnetic activity. Full disk Ca II K images were inspected. (Ca II K images from the National Solar Observatory at Sacramento Peak were used to look for solar faculae and plage near the limb. The Sun was very quiet during both MDI observations with only a single small plage region visible over the entire disk in both 1996 and 1997.) These images were obtained during a 3-day period around the time of the MDI data acquisition and revealed only quiet photosphere near the limb (except for a small plage region near position angle 20 degrees in 1996). The solid line shows our derived full-resolution limb displacement in milliarcsec; the dashed line shows these data boxcar smoothed over a 22-degree interval (and offset by 15 milliarcsec); the dotted line shows the instrumental signal, $I_s(\theta)$, scaled by 0.05; the thick solid line shows the derived limb shape from the 3 term Legendre polynomial fit. Angles of 0, 90, 180 and 270 degrees correspond to the west, north, east and south limbs.

ness signal (which is dominated by higher angular overtones). In addition, the image-stabilization system reduced the spacecraft vibration noise observed in the 1997 measurements by over an order of magnitude.

Limb shape and brightness information is derived from a 6-pixel-wide annulus which contains the solar limb of the MDI continuum data. All of these pixel data are used to derive a mean limb darkening function (LDF) $L(r)$ for each image. The annular data are then divided into $i = 1, \dots, 512$ angular bins, $2\pi/512$ radians wide, with $j = 1, \dots, N_i$ data values ($N_i \approx 35$ pixels) in each bin centred at angle θ_i . Let $d_i(j)$ be the continuum data for pixel j in bin i , where $r_i(j)$ is the distance of pixel j to the centre of the image. Then the average limb displacement at θ_i is determined by a first-order least-squares fit of the data from bin i , $d_i(j)$, to the spatially shifted empirical function, $L(r_i(j) - \beta_i)$ (varying β_i). Thus, the relative solar limb position (β_i) is derived for each of 512 position angles around the limb. The local limb 'brightness' is derived from a least-squares fit of the residual brightness $d_i(j) - L(r_i(j) - \beta_i)$ to $\alpha_i L(r_i(j) - \beta_i)$ (varying α_i). Thus, the α_i parameters describe a least-squares estimate of the local intensity scaling factor by which a given limb angle LDF is 'brighter'.

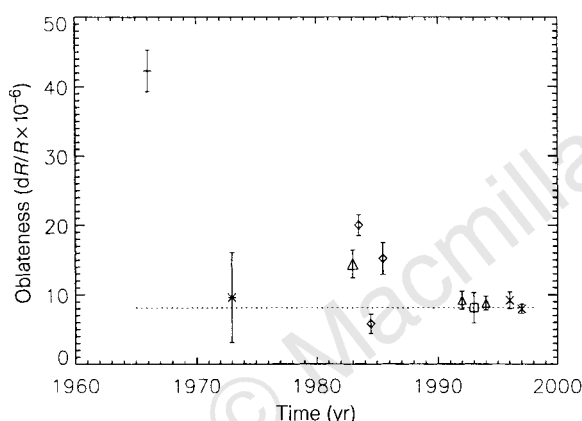
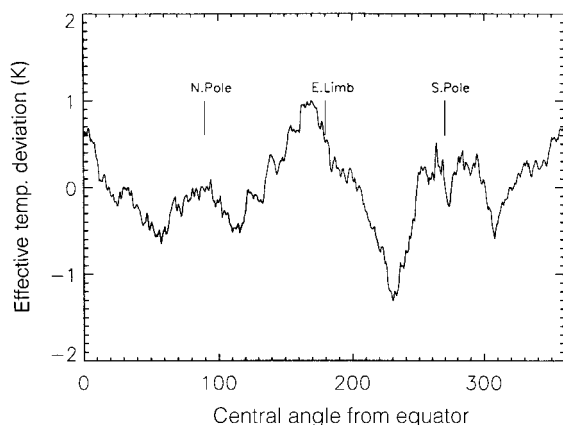


Figure 2 The modern oblateness measurements versus time. These measurements are plotted here in terms of the fractional difference, ϵ , between the equatorial and polar solar radii, where $\epsilon = (r_{\text{equ}} - r_{\text{pole}})/R_{\odot}$ (the apparent solar radius was $R_{\odot} = 972''$ during the MDI measurements). Although there is no evidence of an 11- or 22-year solar cycle trend, it is clear that the measurements and error limits have converged over the past decade towards the flattening implied by the surface photospheric solar rotation rate (as indicated by the dotted line). Symbols indicate the source reference for the results plotted: dash, ref. 5; asterisk, ref. 6; diamonds, ref. 7; large triangle, ref. 8; square, ref. 9; triangles, ref. 10; cross, this work.



The α_i and β_i parameters approximately measure local limb brightness and displacement. This interpretation is approximate because these coefficients are not physically or computationally orthogonal. Furthermore, the limb darkening function may change shape in a way that is more complicated than that described by our parameters, and in response to physical changes in the outer solar atmosphere. Lites¹⁴ has computed specific LDFs for a variety of perturbations. For example, changes in the solar granulation convection, or the response of the atmosphere to the 5-min oscillations, can be reasonably approximated as a scaling and shift of the mean LDF. The detailed physical interpretation of small LDF variations is a fundamental concern for all past and present limb observations and the detection of both limb brightness and shape changes indicates that a more detailed investigation of the LDF variations is warranted.

For MDI, the limiting astrometric statistical noise source is small-scale photometric gain variations caused by the detector and optics—accurate differential gain ('flat-field') calibration of the data is essential. We used a least-squares technique to derive this calibration that involves displacing the telescope and image¹⁵. The MDI flat-field used in this analysis was computed from observations taken in February 1996 and achieved a relative pixel–pixel photometric calibration which is accurate to better than 1%.

The analysis of each satellite roll sequence is performed separately and we label $\alpha_k(\theta_i)$ and $\beta_k(\theta_i)$ as the coefficients derived from roll position, k . Note that the instrumental optical distortion and photometric uncertainties have amplitudes of the order of 0.1 pixel (one pixel subtends $1.96''$) and 1%, and the solar contributions are much smaller than these terms. As the detector (and our reference coordinate frame) rotates with the instrument whereas the solar contribution does not, the data from a given spacecraft roll angle, ω_k , is well described by $\alpha_k(\theta_i) = B(\theta_i) + I_b(\theta_i - \omega_k)$ and $\beta_k(\theta_i) = S(\theta_i) + I_s(\theta_i - \omega_k)$, where $B(\theta)$ and $S(\theta)$ are the solar contributions to the brightness and shape signals at limb angle θ , and I_b and I_s are the corresponding instrumental contributions. Our Sun-centred coordinate frame is orientated so that $\theta = 0$ corresponds to the west equatorial limb, with positive angles increasing towards the solar north pole. In Fig. 1 we show how an algorithm, similar to phase-sensitive detection techniques used in time-series analysis, allows the separation of the low-order harmonics of the solar signals ($B(\theta)$ and $S(\theta)$) from the instrumental noise contributions (I_s and I_b).

Fitting the 1997 $S(\theta)$ results to the above Legendre expansion gives $c_2 = -5.18 \pm 0.44$, and $c_4 = 1.37 \pm 0.54$ milliarcsec. Here we have assumed that the distortion is aligned with the rotation axis to correct for the 6° inclination of the solar axis out of the image plane (by scaling c_2 by 1.02 and c_4 by 1.05). From the 'F-test', we must conclude that the P_4 term is well constrained by the observations as it significantly improves the reduced χ^2_r from the

Figure 3 The effective limb temperature change versus position angle, θ , in 1997 is plotted here. The maximum limb brightness occurs near the equator and poles, with a temperature minimum occurring near 50–60 degrees north and south latitude (that is, near position angles 55, 125, 235 and 305). The form of this latitudinal variation is nearly the same in both hemispheres, although the amplitude of the modulation is larger in the southern hemisphere (about 2.2 K versus 1.4 K in the north). A pronounced 0.5 K brightness dip within a few degrees of the south pole is also apparent in the data. It is notable that the temperature dip is not visible at the north pole, because the north pole was inclined 6° out of the image plane, away from the Earth.

oblateness-only fit at better than the 99% confidence level. Lydon and Sofia⁹ also attempted to measure this shape term in 1992 and 1994 and obtained -0.40 ± 0.44 milliarcsec—consistent with zero, and significantly different from the $l = 4$ term we derive. Although the SOHO data yield the first evidence for a non-zero solar hexadecapole moment, the earlier measurements suggest that the Sun's high-order shape may change in time.

The 1996 data give consistent but noisier results, $c_2 = -6.00 \pm 0.80$ milliarcsec (in 1996, c_4 was indeterminate). The weighted average of our 1996 and 1997 data yield $c_2 = -5.38 \pm 0.39$ milliarcsec (the apparent solar radius was $972''$). Figure 2 summarizes the status of many of the oblateness observations from the past 35 years. In contrast to the $l = 4$ shape measurements, the new MDI data, combined with the most recent results from refs 9 and 10, rule out any 11-year variability in c_2 larger than about 0.5 milliarcsec.

Figure 3 shows that the limb surface brightness, $B(\theta)$, varies with solar latitude. We express this fluctuation in terms of a photometric colour temperature variation by scaling $B(\theta)$ by $T_{\odot}/4$ (taking $T_{\odot} = 5,700$ K for the effective colour temperature). The Princeton–Michigan State–Caltech oblateness/limb photometer had the required accuracy to measure comparably small limb temperature changes, but averaged in relatively larger diametrically opposed limb regions. That experiment² also found a minimum in the solar limb temperature near ± 55 degrees latitude. The MDI brightness profile obtained here is consistent with those earlier results.

The existence of a surface brightness variation (like the polar brightening in Fig. 3) that is not directly associated with photospheric magnetic fields is characteristic of most global models of the solar differential rotation. However, these data show that both the solar shape and brightness are more complicated than a simple quadrupole. It remains to be seen, in detail, what the implications are for our understanding of stellar convection, differential rotation and the solar cycle. If the tiny brightness variations we now see at the surface do 'shadow' the deep convection zone boundary (compare with ref. 2), then we can expect to detect a change in this temperature structure as the solar cycle evolves. Future MDI/SOHO observations should be able to confirm or refute this view. $\square$

Received 22 September; accepted 8 December 1997.

1. Durney, B. R. & Roxburgh, I. Homogeneous convection and the equatorial acceleration. *Sol. Phys.* **16**, 2–20 (1971).
2. Kuhn, J. R. *The Structure of the Sun: VI. Winter School at IAC* (eds Roca Cortes, T. & Sanchez, F.) (Cambridge Univ. Press, 1996).
3. Kuhn, J. R., Libbrecht, K. G. & Dicke, R. H. The surface temperature of the Sun and changes in the temperature constant. *Science* **242**, 908–911 (1988).
4. Schur, W. & Ambronn, L. Messungen des Sonnendurchmessers. *Astron. Mit. Königlichen Stern. Zu Göttingen* **1**–126 (1905).
5. Dicke, R. H. & Goldenberg, H. M. Solar oblateness and general relativity. *Phys. Rev. Lett.* **18**, 313–316 (1967).
6. Hill, H. A. & Stebbins, R. T. The intrinsic visual oblateness of the Sun. *Astrophys. J.* **200**, 471–483 (1987).
7. Dicke, R. H., Kuhn, J. R. & Libbrecht, K. G. Is the solar oblateness variable? *Astrophys. J.* **318**, 471–483 (1975).
8. Beardsley, B. *The Visual Shape and Multipole Moments of the Sun*. Thesis, Univ. Arizona (1987).
9. Rösch, J., Rozelot, J. P., Deslandes, H. & Desnoux, V. A new estimate of the quadrupole moment of the Sun. *Solar Phys.* **165**, 1–11 (1996).
10. Lydon, T. J. & Sofia, S. A measurement of the shape of the solar disk. *Phys. Rev. Lett.* **76**, 177–179 (1996).
11. Rozelot, J. P. & Rösch, J. An upper bound to the solar oblateness. *Solar Phys.* **172**, 11–18 (1997).
12. Scherrer, P. H. et al. *The Solar Oscillations Investigation—Michelson Doppler Imager: The SOHO Mission* (eds Fleck, B., Domingo, V. & Pollard, A.) (Kluwer, Holland, 1995).
13. Kuhn, J. R. et al. *Sounding Solar and Stellar Interiors* (ed. Berthomieu, G.) (Kluwer, Holland, 1997).
14. Lites, B. The extreme limb of the Sun. *Sol. Phys.* **85**, 193–210 (1983).
15. Kuhn, J. R., Lin, H. & Loran, D. Gain calibrating nonuniform image-array data using only the image data. *Publ. Astron. Soc. Pacif.* **103**, 1097–1108 (1991).

Acknowledgements. We thank J. Saba, C. DeForest and J. Covington for their assistance in operating the MDI instrument; L. Allen, H. Schweitzer, J.-P. Olive and K. Miller for the planning and execution of SOHO roll manoeuvres; and the SOHO flight operations team, including H. Benefield, N. Piston and B. Sapper, for making these activities run flawlessly. This work was supported by NASA. We dedicate this work to the memory of Robert H. Dicke, who inspired a generation of astrophysicists to consider the measurement and interpretation of the shape of the Sun.

Correspondence and requests for materials should be addressed to J.R.K. (e-mail address: jkuhn@solar.stanford.edu).

Destruction of the Fermi surface in underdoped high- T_c superconductors

M. R. Norman*, H. Ding*, M. Randeria†, J. C. Campuzano*, T. Yokoya‡, T. Takeuchi§, T. Takahashi§, T. Mochiku¶, K. Kadowaki#, P. Guptasarma* & D. G. Hinks*

* Materials Science Division, Argonne National Laboratory, Argonne, Illinois 60439, USA

† Department of Physics, University of Illinois at Chicago, Chicago, Illinois 60607, USA

‡ Tata Institute of Fundamental Research, Mumbai 400005, India

§ Department of Physics, Tohoku University, 980 Sendai, Japan

¶ Department of Crystalline Materials Science, Nagoya University, Nagoya 464-01, Japan

National Research Institute for Metals, Sengen, Tsukuba, Ibaraki 305, Japan

* Institute of Materials Science, University of Tsukuba, Ibaraki 305, Japan

The Fermi surface—the set of points in momentum space describing gapless electronic excitations—is a central concept in the theory of metals. In this context, the normal 'metallic' state of the optimally doped high-temperature superconductors is not very unusual: above the superconducting transition temperature, T_c , there is evidence for a large Fermi surface^{1–3} despite the absence of well-defined elementary excitations. In contrast, the normal state of underdoped high-temperature superconductors differs in that there is evidence for a 'pseudogap' above T_c (refs 4–6). Here we examine, using angle-resolved photoemission spectroscopy, the temperature dependence of the Fermi surface in underdoped $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$. We find that, on cooling the sample, the pseudogap opens up at different temperatures for different points in momentum space. This leads to an initial breakup of the Fermi surface, at a temperature T^* , into disconnected arcs, which then shrink with decreasing temperature before collapsing to the point nodes of the superconducting ground state below T_c . This unusual behaviour, where the Fermi surface does not form a continuous contour in momentum space as in conventional metals, is unprecedented in that it occurs in the absence of long-range order. Moreover, although the superconducting gap below T_c evolves smoothly into the pseudogap above T_c , the pseudogap differs in its unusual temperature-dependent anisotropy, implying an intimate but non-trivial relationship between the pseudogap and the superconducting gap.

Angle-resolved photoemission spectroscopy (ARPES) probes the occupied part of the electron spectrum, and for quasi-two-dimensional systems its intensity $I(\mathbf{k}, \omega)$ as a function of momentum $\mathbf{k}$ and frequency ω is proportional to the Fermi function $f(\omega)$ times the one-electron spectral function $A(\mathbf{k}, \omega)$ (ref. 3). In Fig. 1a–c, the solid curves are ARPES spectra for an underdoped '85 K' sample of $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$ (Bi2212) at three $\mathbf{k}$ points a, b and c, respectively, on the Fermi surface (determined above T^*) for various temperatures. (Here the Bi2212 samples are distinguished by their onset T_c in this case, 85 K.) We look first at the superconducting state data at $T = 14$ K. At each $\mathbf{k}$ point, the sample spectra are pushed back to positive binding energy ($\omega < 0$) due to the superconducting gap, and we also see a resolution-limited peak associated with a well-defined quasiparticle excitation in the superconducting state. The superconducting gap, as estimated by the position of the midpoint of the leading edge of the spectrum, is seen to decrease as one moves from point 'a' near $\bar{M}$ to 'b' to 'c', closer to the diagonal $\Gamma - Y$ direction (notation is described in Fig. 1 legend), consistent with a $d_{x^2-y^2}$ order parameter. Next, we consider the changes in Fig. 1 as a function of increasing T . At each $\mathbf{k}$ point the quasiparticle

peak disappears above T_c , but the suppression of spectral weight—the pseudogap—persists well above T_c , as noted in earlier work^{4–6}.

The striking new feature which is apparent from Fig. 1 is that the pseudogap at different k points closes at different temperatures, with larger gaps persisting to higher T s. At point a, near $\bar{M}$, there is a pseudogap at all T s below 180 K, at which the Bi2212 leading edge matches that of Pt. We take this as the definition of T^* (ref. 5) above which the largest pseudogap has vanished within the resolution of our experiment, and a closed contour of gapless excitations—a Fermi surface—is obtained⁷. The surprise is that if we move along this Fermi surface to point b the sample leading edge matches Pt at 120 K, which is smaller than T^* . Continuing to point c, about halfway to the diagonal direction, we find that the Bi2212 and Pt leading edges match at an even lower temperature of 95 K. In addition, we have measured spectra on the same sample along the Fermi contour near the ΓY line and found no gap at any T , even below T_c , consistent with $d_{x^2-y^2}$ anisotropy.

One simple way to quantify the behaviour of the gap is to plot the midpoint of the leading edge of the spectrum (Fig. 1e). We will say that the pseudogap has closed at a k point when the midpoint equals zero energy, in accordance with the discussion above. From this plot, we find that the pseudogap closes at point a at a T above 180 K,

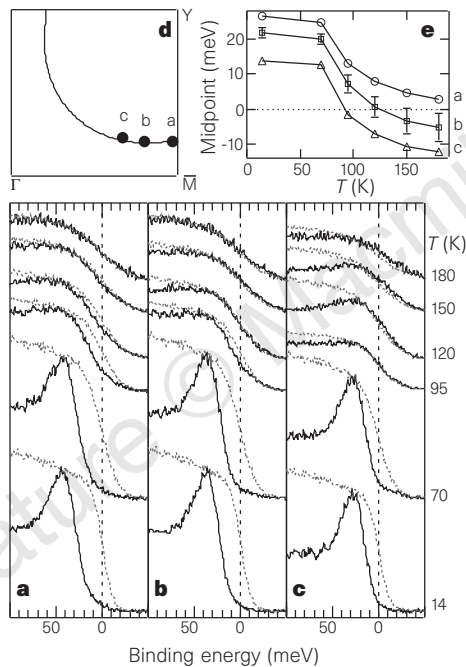


Figure 1 Data obtained on single crystals of Bi2212 grown by the travelling solvent floating-zone method. Doping was achieved by adjusting the oxygen partial pressure during annealing with samples labelled by their onset T_c s. Measurements were carried out at the Synchrotron Radiation Center, Wisconsin, using a high resolution 4-m normal incidence monochromator with 22-eV photons and an energy resolution of 20 meV (full-width at half-maximum). The spectra in **a–c** are taken at three k points in the Brillouin zone, shown in **d**, for an 85 K underdoped Bi2212 sample at various temperatures (solid curves). (The Y quadrant was studied to minimize effects due to the superlattice¹².) Our notation is $\Gamma = (0, 0)$, $\bar{M} = (\pi, 0)$ and $Y = (\pi, \pi)$, in units of $1/a$, where a is the Cu–Cu distance, and $\Gamma\bar{M}$ is along the CuO bond direction. The dotted curves are reference spectra from polycrystalline Pt (in electrical contact with the sample) used to determine the chemical potential (zero binding energy). Note the closing of the spectral gap at different T for different k . This feature is also apparent in the plot (**e**) of the midpoint of the leading edge of the spectra as a function of T .

at point b at 120 K, and at point c just below 95 K. If we now view these data as a function of decreasing T , the picture of Fig. 2 clearly emerges. The pseudogap suppression first opens up near $(\pi, 0)$ and progressively removes larger portions of the Fermi contour, leading to gapless arcs which shrink with decreasing T . We note that midpoints with negative binding energy, particularly for k point c, indicate the formation of a peak in the spectral function at $\omega = 0$ as T increases.

We see similar results on other underdoped samples of Bi2212. For example, in the upper panel of Fig. 3 we show midpoints for a 77 K underdoped sample at two k points shown in the inset, with behaviour very similar to that of the 85 K sample of Fig. 1. This behaviour may be compared with that of the more conventional T -dependence of an overdoped 87 K sample as shown in the lower panel. Gaps with different magnitudes, one at a k point near $\bar{M}$ and the other halfway towards the ΓY direction, go to zero at the same temperature, very close to T_c ; we have also seen this in other overdoped samples. This is in marked contrast with the new results on underdoped samples. Further, to show that the negative midpoints at high T s are not unusual, we plot those for an 82 K overdoped sample at the $\bar{M}Y$ Fermi point as filled symbols in the lower panel of Fig. 3. The midpoint goes to zero at about T_c

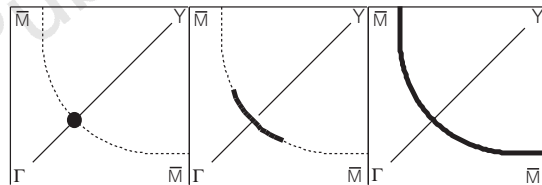


Figure 2 Schematic illustration of the temperature evolution of the Fermi surface in underdoped copper oxides. The d -wave node below T_c (left panel) becomes a gapless arc above T_c (middle panel) which expands with increasing T to form the full Fermi surface at T^* (right panel).

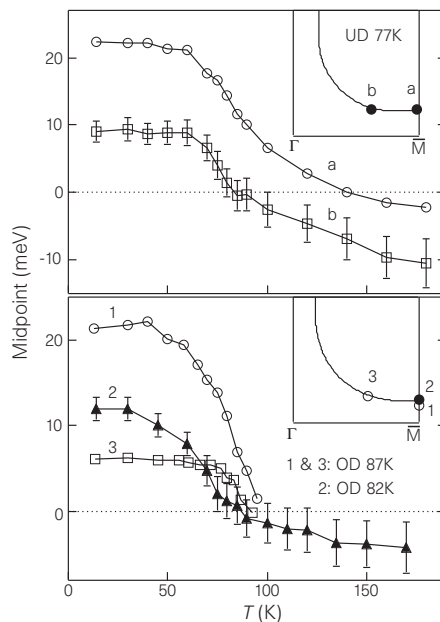


Figure 3 Midpoints of the leading edge of the ARPES spectra of Bi2212 samples, plotted against temperature. Top panel, data for a 77 K underdoped (UD) sample, again showing closure of the spectral gap at different T for different k . This behaviour can be contrasted with that of overdoped samples (bottom panel) where all gaps close near T_c .

(indicating the absence of a pseudogap above T_c in this sample) followed by a slower evolution to negative binding energy (indicating the formation of a spectral peak, as discussed above). There are also important differences between the spectral lineshapes of the overdoped and underdoped cases. The underdoped spectra are broader above T_c , as demonstrated by the flatness of the spectra seen in Fig. 1 at $\mathbf{k}$ points a and b, and have smaller quasiparticle peaks below T_c , implying an increase in the strength of the interactions as the doping is reduced.

Before discussing the implications of our results, we introduce a visualization aid that makes these results easier to understand. This symmetrization method, described in Fig. 4 legend, effectively eliminates the Fermi function f from ARPES data and permits us to focus directly on the spectral function A . We have extensively checked this method, and have studied in detail the errors introduced by incorrect determination of either the chemical potential or the Fermi momentum $\mathbf{k}_F$ (which lead to spurious narrow features in the symmetrized spectra); we have also studied the effect of the small (1° radius) $\mathbf{k}$ -window of the experiment (this effect was found to be small).

In Fig. 4a–c we show symmetrized spectra for the 85 K underdoped sample corresponding to the raw data of Fig. 1a–c, respectively. To emphasize that the symmetry is put in by hand, we show the $\omega > 0$ curve as a dotted line. In Fig. 4a at $\mathbf{k}$ point a near $\bar{M}$, the sharp quasiparticle peak disappears above T_c but a strong pseudogap suppression, on the same scale as the superconducting gap, persists all the way up to 180 K (T^*). In Fig. 4b and c we again see pseudogap depressions on the scale of the superconducting gaps at those points, but the pseudogap fills up at lower temperatures: 120 K at b and 95 K at c. In Fig. 4c, moreover, a spectral peak at zero energy emerges as T is raised. All of the conclusions drawn from the

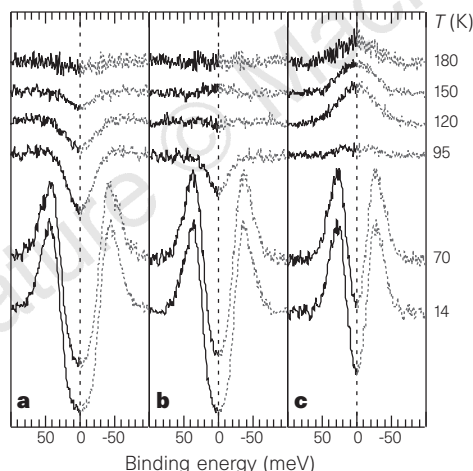


Figure 4 Results of symmetrization analyses. Given ARPES data described³ by $I(\omega) = \sum_{\mathbf{k}} f(\omega) A(\mathbf{k}, \omega)$ (with the sum over a small momentum window about the Fermi momentum $\mathbf{k}_F$), we can generate the symmetrized spectrum $I(\omega) + I(-\omega)$. Making the reasonable assumption of particle-hole (p-h) symmetry for a small range of ω and $\epsilon_{\mathbf{k}}$, we have $A(\epsilon_{\mathbf{k}}, \omega) = A(-\epsilon_{\mathbf{k}}, -\omega)$ for $|\omega|, |\epsilon_{\mathbf{k}}|$ less than few tens of meV (where $\epsilon_{\mathbf{k}}$ are the bare energies). It then follows, using the identity $f(-\omega) = 1 - f(\omega)$, that $I(\omega) + I(-\omega) = \sum_{\mathbf{k}} f(\omega) A(\mathbf{k}, \omega)$ which is true even after convolution with a (symmetric) energy resolution function. This symmetrized spectrum coincides with the raw data for $\omega \leq -2.2T_{\text{eff}}$, where $4.4T_{\text{eff}}$ is the 10–90% width of the Pt leading edge, which includes the effects of both temperature and resolution. Non-trivial information is obtained for the range $|\omega| \leq 2.2T_{\text{eff}}$, which is then the scale on which p-h symmetry has to be valid. The curves are symmetrized spectra corresponding to the raw spectra of Fig. 1. The gap closing in the raw spectrum of Fig. 1 corresponds to where the pseudogap depression disappears in the symmetrized spectrum. We note the appearance of a spectral peak at higher temperatures in c.

raw data in Figs 1 and 3 are obvious from the simple symmetrization analysis of Fig. 4.

We now discuss why the T dependence of the Fermi arc is not simply due to inelastic scattering above T_c broadening the d -wave node. From Fig. 4, it is apparent that the gap ‘fills in’ for $\mathbf{k}$ points a and b as T is raised, whereas it ‘closes’ for $\mathbf{k}$ point c as a peak at zero energy emerges. This can be seen more clearly in Fig. 5, where we show symmetrized spectra for a 75 K underdoped sample of Bi2212 at two $\mathbf{k}$ points (similar to points a and c of Fig. 4) as a function of temperature. For the first point (I; top panel), the spectral feature at the gap edge does not move with temperature, whereas for the latter point (II; bottom panel), it clearly moves in to zero energy.

A unique feature of ARPES is that it provides $\mathbf{k}$ -resolved information. We believe that the unusual T -dependence of the pseudogap anisotropy will be a very important factor in reconciling the different crossovers seen in the pseudogap regime by different probes, each of which measures a $\mathbf{k}$ -sum weighted with a different set of $\mathbf{k}$ -dependent matrix elements or kinematical factors (such as Fermi velocity). For instance, quantities which involve the Fermi velocity, such as the d.c. resistivity above T_c and the penetration depth below T_c (superfluid density), should be sensitive to the region near the ΓY direction, and would thus be affected by the behaviour we see at $\mathbf{k}$ point c. Other types of measurements (such as specific heat and tunnelling) are more ‘zone-averaged’ and will have significant contributions from $\mathbf{k}$ points a and b as well, so should see a more pronounced pseudogap effect. Other data that we have indicate that the region in the Brillouin zone where behaviour like $\mathbf{k}$ -point c is seen shrinks as the doping is reduced, and thus appears to be correlated with the loss of superfluid density⁸. Further, we speculate that the disconnected Fermi arcs should have a profound influence on magnetotransport, given the lack of a continuous Fermi contour in momentum space.

We emphasize that the Fermi arcs do not imply the existence of hole pockets (that is, small closed contours) centred about $(\pi/2, \pi/2)$, as suggested by some theories of lightly doped Mott insulators (for a review, see ref. 9). In the samples studied here (and more heavily underdoped ones) we have carefully searched for hole pockets and for shadow band dispersion, but found no evidence for either⁷. The

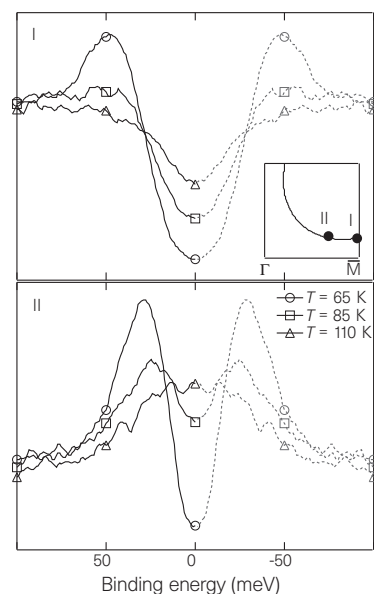


Figure 5 Symmetrized spectra for a 75 K underdoped sample of Bi2212, at three different temperatures. The $\mathbf{k}$ points are analogous to a and c of Fig. 4. We note that the spectral feature at the gap edge does not move in energy with increasing T for point I (top panel), but does move in to zero energy for point II (bottom panel).

gapless arcs that we observe are simply an intermediate state in the smooth evolution of d -wave nodes into a full Fermi surface. This smooth evolution was carefully checked on an 83 K underdoped sample where a detailed sweep was done in $\mathbf{k}$ space at $T = 90$ K, revealing only a small Fermi arc just above T_c . This behaviour is fully consistent with the gap above and below T_c being of the same origin as suggested by our earlier experiments^{5,7}.

Theoretical calculations in which d -wave pairing correlations cause a pseudogap above T_c (ref. 10) have predicted gapless arcs which expand as T increases. Resonating valence bond theories also lead to gapless arcs above T_c due to spinon pairing¹¹. There are other proposals in which the pseudogap has a completely different (non-pairing) origin from the superconducting gap. Given the smooth evolution that we find through T_c , such proposals seem difficult to reconcile with our results. □

Received 23 July; accepted 22 December 1997.

1. Campuzano, J. C. *et al.* The Fermi surfaces of $\text{YBa}_2\text{Cu}_3\text{O}_{6.9}$ as seen by angle-resolved photoemission. *Phys. Rev. Lett.* **64**, 2308–2312 (1990).
2. Olson, C. G. *et al.* High-resolution angle-resolved photoemission study of the Fermi surface and normal-state electronic structure of $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_8$. *Phys. Rev. B* **42**, 381–386 (1990).
3. Randeria, M. *et al.* Momentum distribution sum rule for angle-resolved photoemission. *Phys. Rev. Lett.* **74**, 4951–4954 (1995).
4. Marshall, D. S. *et al.* Unconventional electronic structure evolution with hole doping in $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$: angle-resolved photoemission results. *Phys. Rev. Lett.* **76**, 4841–4844 (1996).
5. Ding, H. *et al.* Spectroscopic evidence for a pseudogap in the normal state of underdoped high- T_c superconductors. *Nature* **382**, 51–54 (1996).
6. Loeser, A. G. *et al.* Excitation gap in the normal state of underdoped $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$. *Science* **273**, 325–329 (1996).
7. Ding, H. *et al.* Evolution of the Fermi surface with carrier concentration in $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$. *Phys. Rev. Lett.* **78**, 2628–2631 (1997).
8. Uemura, Y. J. *et al.* Universal correlations between T_c and n_d/m^* in high- T_c cuprate superconductors. *Phys. Rev. Lett.* **62**, 2317–2320 (1989).
9. Lee, P. A. in *High Temperature Superconductivity* (eds Bedell, K. S. *et al.*) 96–116 (Addison-Wesley, New York, 1990).
10. Engelbrecht, J. R., Nazarenko, A., Randeria, M. & Dagotto, E. Pseudogap above T_c in a model with $d_{x^2-y^2}$ pairing. Preprint available at <http://xxx.lanl.gov/archive/cond-mat/9705166>
11. Wen, X. G. & Lee, P. A. Theory of underdoped cuprates. *Phys. Rev. Lett.* **76**, 503–506 (1996).
12. Ding, H. *et al.* Angle-resolved photoemission spectroscopy study of the superconducting gap anisotropy of $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+x}$. *Phys. Rev. B* **54**, R9678–B9681 (1996).

Acknowledgements. We thank J. Sadleir and A. Kaminski for their help. This work was supported by the US Dept of Energy Basic Energy Sciences, the US NSF, the NSF Science and Technology Center for Superconductivity, the CREST of JST, and the Ministry of Education, Science and Culture of Japan. The Synchrotron Radiation Center is supported by the NSF.

Correspondence and requests for materials should be addressed to J.C.C. (e-mail: jcc@uic.edu).

A nanometre-scale mechanical electrometer

A. N. Cleland* & M. L. Roukes

Condensed Matter Physics 114-36, California Institute of Technology, Pasadena, California 91125, USA

The mechanical detection of charge has a long history, dating back more than 200 years to Coulomb's torsion-balance electrometer¹. The modern analogues of such instruments are semiconductor-based field-effect devices, the most sensitive of which are cryogenically cooled single-electron transistors². But although these latter devices have extremely high charge sensitivity, they suffer from limited bandwidth and must be operated at millikelvin temperatures in order to reduce thermal noise. Here we report the fabrication and characterization of a working nanometre-scale mechanical electrometer. We achieve a charge sensitivity of $0.1 e \text{ Hz}^{-0.5}$, competitive with conventional semiconductor field-effect transistors; moreover, thermal noise analysis indicates that the nanometre-scale electrometer should ultimately reach sensitivities of the order of $10^{-6} e \text{ Hz}^{-0.5}$, comparable with charge-detection capabilities of cryogenic single-electron transistors. The nanometre-scale electrometer has the additional advantages of

high temperature (≥ 4.2 K) operation and response over a larger bandwidth, from which a diversity of applications may result.

We have described elsewhere³ the techniques used to fabricate and measure devices of the type reported here. An electron micrograph and a schematic view of our nanomechanical electrometer is shown in Fig. 1. It consists of three principal components: electrodes, which experience an attractive force when a small charge is applied; a compliant mechanical element that moves in response to this force; and a displacement detector that provides a means of monitoring the motion. The mechanical element is a two-element torsional resonator with a spring constant G_0 and moment I . Its fundamental torsional resonance frequency is ω_0 , and its mechanical loss is parametrized by a quality factor $Q = \omega_0/\Delta\omega$, where $\Delta\omega$ is the frequency width of the resonant response at half maximum.

The device includes three electrodes: two for inducing and measuring the mechanical response of the structure, and one for coupling charge, which alters this response. Two electrodes consist of metal loops tracing the boundaries of the outer and inner paddles, while the opposing metal gate electrode is fixed to the stationary substrate, at a distance d from the inner paddle. The mutual capacitance between the gate electrode and the inner paddle is represented by the parameter C . To measure a small charge, the gate electrode is biased by a charge q_0 , which yields an electrostatic force $F_E = q_0^2/Cd$. Small changes in the coupled charge, δq , about the bias point q_0 alter the force by an amount $f_E = 2\delta q E_0$, where E_0 is the equilibrium electric field. This results in an effective torsional spring constant $G_{\text{eff}} = G_0 + g$, where $g = -r\partial q\partial E_0/\partial\theta$ (θ is the paddle torsion angle, E_0 the field component along the unit vector θ , and r the paddle's radial dimension). This gives rise to a shift away from the unperturbed resonance frequency, $\delta\omega/\omega_0 \approx g/2G_0$.

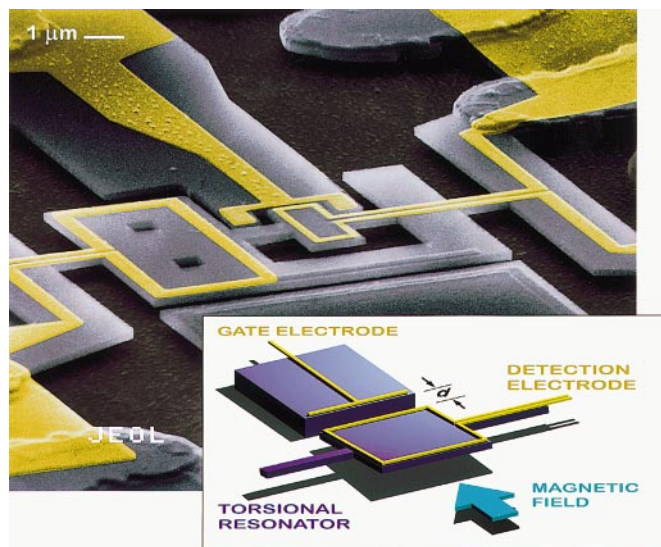


Figure 1 Nanometre-scale charge detector. The inset schematically depicts its principal components: torsional mechanical resonator, detection electrode, and gate electrode used to couple charge to the mechanical element. An external, parallel magnetic field is employed for readout. In the actual device (main figure) a double torsional element is used with moment I , torsional spring constant G_0 (calculated to be $1.1 \times 10^{-10} \text{ N m}$), and spacing, $d = 0.5 \mu\text{m}$, between the gate and inner paddle electrodes. The fundamental resonance frequency of the structure is 2.61 MHz, with a quality factor measured to be $Q = 6,500$; the coupling capacitance between the gate and resonator was calculated to be $C = 0.4 \text{ fF}$ (H. Pothier, D. Esteve and M. H. Devoret, personal communication). The resonator was fabricated from a single-crystal Si-on-insulator substrate, with a $0.2\text{-}\mu\text{m}$ -thick Si layer on a $0.4\text{-}\mu\text{m}$ insulating layer. The Au electrodes and resonator structure are patterned using electron beam lithography. The smallest lateral feature in the structure is $0.2 \mu\text{m}$, so there is scope to scale these devices to even smaller dimensions. This could ultimately yield bandwidths in the GHz regime.

* Present address: Department of Physics, UC Santa Barbara, Santa Barbara, California 93106, USA.

To perform electrometry we monitor either the charge-induced shift in the rest point of the torsional resonator, or the change in the resonance frequency due to the charge modulation of G_{eff} . The latter approach can be especially sensitive if the quality factor of the resonator is high. We have employed frequency modulation detection⁴ to obtain enhanced sensitivity and large measurement bandwidth.

When optimally configured, thermally induced motion of the resonator limits the mechanical electrometer's sensitivity. Additional noise may be introduced by the displacement detector: with conventional optical detection this arises from photon shot noise, whereas with electromagnetic detection (used here) this arises from the readout amplifier. We parametrize this additional noise by a detection noise temperature T_N , while representing the thermal fluctuations of the mechanical resonator by a temperature T_R . For a paddle driven with a motional amplitude θ_{max} , the spectral density of the frequency noise is given as⁵:

$$S_\omega = \frac{\omega_0 k_B (T_R + T_N)}{Q G_0 \theta_{\text{max}}^2} \quad (1)$$

The corresponding charge noise spectral density is given by:

$$S_q = \frac{G_0}{\partial E_0 / \partial \theta} \frac{1}{\omega_0^2 S_\omega} \quad (2)$$

In our experiments we use these devices for charge detection by employing the fundamental torsional mode (both elements of the torsional structure vibrate with similar amplitude and phase). The resonator was placed in vacuum at 4.2 K. A radio frequency (r.f.) current was passed through the larger gold loop which, in the presence of an applied 8 T magnetic field, generates a r.f. torque along the support axis of the resonator. The resulting torsional motion of the resonator's inner paddle in the applied field generates an electromotive force (EMF) that is subsequently amplified and detected by room-temperature electronics. Related techniques have been used to detect the motion of flexural resonators^{6,7}.

In Fig. 2 we show the amplitude of the EMF measured as a

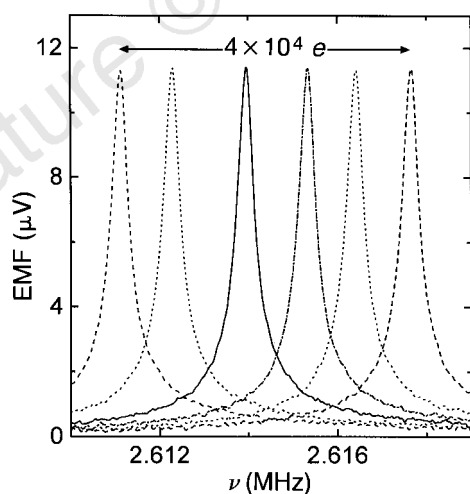


Figure 2 Amplitude of the induced EMF across the small loop (inner paddle) of the resonator, as a function of the frequency of the drive current passed through the large loop (outer paddle). The charge coupled by the gate electrode was changed in steps for each trace, with a total range of $4 \times 10^4 e$ in the coupled charge. The charge sensitivity is referenced to the charge coupled directly to the resonator paddle; we ignore the large amount of charge induced by our voltage bias on the parasitic capacitances associated with the cabling. In real charge-detection applications of this device, the charge source would necessarily be placed, at most, a few tens of micrometres from the gate. This constraint is identical to those arising in applications based on the single-electron transistor, and is not a serious limitation.

function of drive current frequency, with fixed current amplitude. Each trace was taken with a different amount of d.c. voltage applied to the gate electrode. The range of voltages applied corresponds to a total change in the coupled charge of $\sim 4 \times 10^4 e$.

A more sensitive detection technique is to fix the frequency of the drive current at the mechanical resonance frequency, and monitor the phase of the induced EMF relative to the phase of the drive current. The circuit used for this measurement is shown in Fig. 3. In the inset of Fig. 4 we show the frequency-dependent phase responsibility $R = \partial \phi / \partial q(\nu)$, which is the measured variation of the EMF phase as a function of d.c. coupled charge. By measuring the spectral density of the phase noise, $S_\phi(\nu)$, we can then deduce the spectral density of the charge noise $S_Q(\nu) = S_\phi(\nu)/R^2$ (Fig. 4). We find the charge noise to be dominated by $1/f$ noise at low frequencies, with a noise level of $S_Q^{0.5} = 0.6 e \text{ Hz}^{-0.5}$ at $\nu = 10 \text{ Hz}$. Above a frequency of about 500 Hz the noise levels out to a white noise floor of $S_Q^{0.5} = 0.09 e \text{ Hz}^{-0.5}$. The lowest level of charge noise obtained is $\sim 0.1 e \text{ Hz}^{-0.5}$, competitive with state-of-the-art electrometers based on cryogenically cooled field-effect transistors⁸.

Mechanical detection of charge has been previously demonstrated^{9,10} using an atomic force microscope (AFM) apparently providing sensitivity as low as $0.03 e \text{ Hz}^{-0.5}$. This was achieved using cantilevers with $\sim 100 \mu\text{m}$ dimensions and resonance frequencies in the audio range. Their intrinsically low response frequency and the overall size of these macroscopic AFM instruments makes them impractical for general applications in electrometry. Nonetheless, one may consider our devices to be an extension of the principles demonstrated by such AFM charge detection—but carried to much smaller geometries, higher resonance frequencies, and achieved within an entirely integrated, nanomachined structure.

These nanometre-scale mechanical electrometers show that greatly enhanced sensitivity can be expected with future improvements to device design. Our measurements indicate that the $1/f$ noise in these devices emanates from the resonator or the charge coupled to it, similar to findings with the single electron transistor². This should be susceptible to careful surface passivation. The measurement bandwidth can be increased using feedback techniques⁴, with which we achieved a bandwidth of 10 kHz and a noise of $2 e \text{ Hz}^{-0.5}$, limited by the room-temperature feedback amplifier. Use of a cryogenic amplifier would lead to significantly improved sensitivity.

We now consider the ultimate theoretical limits to the performance of these devices. The charge noise floor is set by the precision of the measurement of resonance frequency. This, in turn, is limited

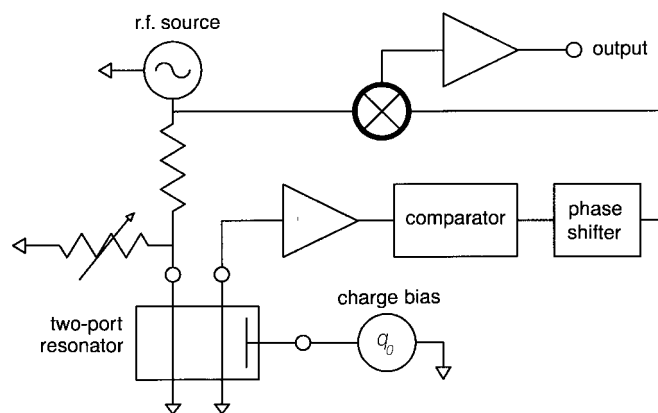


Figure 3 Measurement scheme used to measure changes in the phase of the induced EMF as a function of coupled charge, for fixed drive current frequency of 2.617 MHz. The bandwidth obtained with this technique is limited to $\nu_0/Q = 400 \text{ Hz}$, due to uncontrolled parasitic capacitances in these prototype devices.

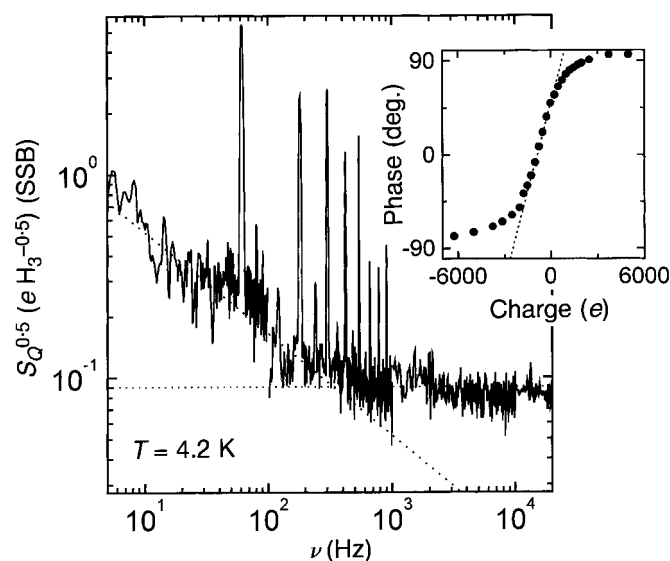


Figure 4 Measured charge noise as a function of frequency, using the measurement set-up shown in Fig. 3. The noise, measured at the output of a mixer, is scaled to single-sideband value. The horizontal dotted line is the white noise level of $0.09 \text{ e Hz}^{-0.5}$, and the sloped dotted line is the level of $1/f$ noise. Inset, measured phase variation as a function of d.c. coupled charge to the gate. The dotted line indicates the zero-frequency charge responsivity. Measurement performed using set-up in Fig. 3.

by thermal noise in the resonator and the readout amplifier, as given by equation (1). For enhanced sensitivity, we need low noise temperatures, high quality factors, Q , and large drive amplitudes, θ (limited, however, by the onset of nonlinearity at large θ). Large-radii paddles with small electrode spacing are advantageous, but the accompanying increase in the paddle inertia I , and reduction in frequency ν_0 , is deleterious, so careful optimization is required.

For our device design we can estimate the thermal limit to the charge sensitivity. With a drive amplitude of 30 mrad and a readout amplifier noise temperature of 300 K (roughly the value in the experiment), a voltage bias of 10 V on the gate gives a thermally limited charge noise of $7 \times 10^{-4} \text{ e Hz}^{-0.5}$. A state-of-the-art cryogenic amplifier (noise temperature $\sim 10 \text{ K}$) reduces this to $1 \times 10^{-4} \text{ e Hz}^{-0.5}$. This becomes comparable to a single-electron transistor operating at 50 mK and measured at 1 kHz, where it is limited by $1/f$ process^{2,11}. If we reduce the torsional rod cross-section while increasing its length, so that each rod has dimensions $0.1 \times 0.1 \times 2 \mu\text{m}^3$, and the paddle dimensions are reduced to $0.1 \times 0.5 \times 1 \mu\text{m}^3$ (here the last dimension is the paddle radius, r), we obtain a resonance frequency $\nu_0 = 7 \text{ MHz}$. Placing the gate electrode $0.1 \mu\text{m}$ from the paddle edge, we obtain a charge noise floor of $\sim 3 \times 10^{-6} \text{ e Hz}^{-0.5}$, an improvement of nearly two orders of magnitude.

These experimental prototypes clearly show that nanometre-scale mechanical electrometry provides a new means for ultrasensitive charge measurement; they are unique in that they offer a much higher potential bandwidth than the single-electron transistor. A further advantage is their ability to operate at 4.2 K (or higher), as compared to the millikelvin temperatures required for very low noise single-electron transistors. This opens possibilities for applications such as single-photon photodetection¹² or ultrasensitive scanned electrometry¹³. □

Received 14 August; accepted 29 December 1997.

1. Coulomb, C. A. *Memoires de l'Academie Royale des Sciences* 229–236 (Academie royale des sciences, Paris, 1784).
2. Zimmerli, G., Eiles, T. M., Kautz, R. L. & Martinis, J. M. Noise in the Coulomb blockade electrometer. *Appl. Phys. Lett.* **61**, 237–239 (1992).

3. Cleland, A. N. & Roukes, M. L. Fabrication of high frequency nanometer scale mechanical resonators from bulk Si crystals. *Appl. Phys. Lett.* **69**, 2653–2655 (1996).
4. Albrecht, T. R., Grutter, P., Horne, D. & Rugar, D. Frequency modulation detection using high- Q cantilevers for enhanced force microscope sensitivity. *J. Appl. Phys.* **69**, 668–673 (1991).
5. Robins, W. P. *Phase Noise in Signal Sources: Theory and Applications* (Peregrinus, London, 1984).
6. Greywall, D. S., Yurke, B., Busch, P. A., Pargellis, A. N. & Willett, R. L. Evading amplifier noise in nonlinear oscillators. *Phys. Rev. Lett.* **72**, 2992–2995 (1994).
7. Yurke, B., Greywall, D. S., Pargellis, A. N. & Busch, P. A. Theory of amplifier noise evasion in an oscillator employing a nonlinear resonator. *Phys. Rev. A* **51**, 4211–4229 (1995).
8. Bordoni, F., Maggi, G., Ottaviano, A. & Pallotino, G. V. Very low noise cooled audiofrequency preamplifier for gravitational research. *Rev. Sci. Instrum.* **52**, 1079–1086 (1981).
9. Stern, J. E., Terris, B. D., Mamin, H. J. & Rugar, D. Deposition and imaging of localized charge in insulator surfaces using a force microscope. *Appl. Phys. Lett.* **53**, 2717–2719 (1988).
10. Schonenberger, C. & Alvarado, S. F. Observation of single charge carriers by force microscopy. *Phys. Rev. Lett.* **65**, 3162–3164 (1990).
11. Pendry, J. B., Kirkman, P. D. & Castano, E. Electrons at disordered surfaces and $1/f$ noise. *Phys. Rev. Lett.* **57**, 2983–2986 (1986).
12. Cleland, A. N., Esteve, D., Urbina, C. & Devoret, M. H. Very low noise photodetector based on the single electron transistor. *Appl. Phys. Lett.* **61**, 2820–2822 (1992).
13. Yoo, M. J. *et al.* Scanning single-electron transistor microscopy: Imaging individual charges. *Science* **276**, 579–582 (1997).

Acknowledgements. This work was supported by DARPA ETO/MEMS.

Correspondence and requests for materials should be addressed to M.L.R. (e-mail: roukes@caltech.edu).

Discovery of a useful thin-film dielectric using a composition-spread approach

R. B. van Dover, L. F. Schneemeyer & R. M. Fleming

Bell Laboratories, Lucent Technologies, Murray Hill, New Jersey 07974, USA

The continuing drive towards miniaturization of electronic devices¹ is motivating the search for new materials. Consider, for example, the case of the much-used dynamic random-access memory. The minimum capacitance per cell that can be tolerated is expected² to remain at 30–40 fF, but as the cell area decreases, the corresponding reduction in geometric capacitance has to be compensated for. So far, this has been achieved by resorting to complex non-planar structures and/or using much thinner films of the dielectric insulator, amorphous silicon dioxide (a-SiO_x), although the latter approach is limited by the electric fields that can be supported by a-SiO_x before its insulating properties break down. An alternative strategy is to develop thin-film insulators that have a dielectric constant significantly greater than that of a-SiO_x , reducing the size of the fields required for device operation. Here we show that a composition-spread technique allows for the efficient evaluating of materials with both a high dielectric constant and a high breakdown field. We apply this approach to the Zr–Sn–Ti–O system, and we find that compositions close to $\text{Zr}_{0.15}\text{Sn}_{0.3}\text{Ti}_{0.55}\text{O}_{2-\delta}$ are better thin-film dielectrics than high-quality deposited a-SiO_x . Although detailed tests of the performance of these materials have not yet been carried out, our initial results suggest that they are likely to be comparable to the best alternatives (such as $(\text{Ba}, \text{Sr})\text{TiO}_3$) currently being considered for integrated-circuit capacitors.

Initial assessment of a new material demands the choice of a suitable criterion that can distinguish promising materials from those with little potential. For capacitor materials, the key parameters are the capacitance per unit area (C/A) and maximum working voltage (V_{br}). A useful figure of merit³ (FOM) can be defined as the product of these, that is, $\text{FOM} = CV_{\text{br}}/A$. As $C = \epsilon_r \epsilon_0 A/d$ and $V_{\text{br}} = E_{\text{br}}d$, where d is the dielectric thickness, ϵ_r the dielectric constant, ϵ_0 the permittivity of free space, and E_{br} the breakdown field, FOM can also be defined as $\text{FOM} = \epsilon_r \epsilon_0 E_{\text{br}}$. Defined in this way, FOM corresponds to the maximum charge that can be stored on a capacitor made from a given material. It is independent of film thickness to the degree that ϵ_r and E_{br} are. Note that the useful working voltage may be much less than the maximum working

voltage if the material does not fail by dielectric breakdown at a given field, but by another degradation mechanism (such as ionic-transport-induced resistance degradation and/or thermal runaway)^{4,5}. Furthermore, in any specific application considerations other than dielectric failure may be more important, such as low deposition temperature, low leakage current, low temperature coefficient, or high compatibility with conventional integrated circuit (IC) processing.

Typical deposited films of standard a-SiO_xN_y have $\epsilon_r = 4.0$ and $E_{br} \sim 10 \text{ MV cm}^{-1}$, for a FOM of $3.5 \mu\text{C cm}^{-2}$. Recent attention has focused on other materials with high values of ϵ_r and FOM. Amorphous tantalum pentoxide, a-Ta₂O_{5-δ}, typically has $\epsilon_r \sim 23$ and $E_{br} \sim 4 \text{ MV cm}^{-1}$, giving FOM $\sim 8.1 \mu\text{C cm}^{-2}$. Polycrystalline x-(Ba, Sr)TiO₃ typically⁶ has $\epsilon_r \sim 200$, $E_{br} = 1 \text{ MV cm}^{-1}$, giving FOM $\sim 18 \mu\text{C cm}^{-2}$, although values as high as $41 \mu\text{C cm}^{-2}$ have been reported⁷ (inferred using the criterion defined below). The high dielectric constant of x-(Ba, Sr)TiO₃ is quite attractive, but a film formation requires a high deposition temperature (typically $>650^\circ\text{C}$) and exotic electrode metallurgy. Furthermore, the dielectric constant of this material depends strongly on the thickness of the film⁸. Amorphous Ta₂O_{5-δ} films exhibit a much smaller thickness dependence, and the material is generally considered compatible with Si IC fabrication technology.

In searching for alternative thin-film dielectrics, we have focused on deposition at substrate temperatures below 300°C to maintain compatibility with back-end Si IC fabrication technology. Low-temperature deposition also produces amorphous films, avoiding defects at grain boundaries that might degrade E_{br} . We have chosen to consider only the elements that are most compatible with Si IC technology—eschewing, for example, late transition metals. To evaluate a wide range of compositions efficiently, we have developed the technique of depositing a single film with a ternary composition spread, and evaluating the critical properties as a function of position (which is directly related to material composition) using an automated tool—the continuous composition spread (CCS) technique. It is related to the discrete combinatorial synthesis (DCS) technique proposed by Xiang and collaborators⁹ in that it employs an automated measurement technique to evaluate many compositions without human interaction. However, the DCS technique involves deposition of precursor layers followed by a high-temperature reaction, and is thus not suitable for exploration of low-temperature-processed materials. Our technique is also related to that used by Sawatzky and Kay¹⁰ (one-dimensional spread) and Hanak¹¹ (two-dimensional spread), differing mainly in the versatility obtained with independently powered magnetron guns and in our use of fully automatic measurements to characterize the properties of the resultant composition-spread films. It is the combination of both multicomponent synthesis and automated measurements that allows the spectacular increase in throughput of materials measured. We typically synthesize and evaluate 4,000 samples in a ternary composition diagram in one day. We have evaluated over thirty multicomponent oxide systems, among which the Zr-Ti-Sn-O system yielded the highest FOM.

Films were deposited by off-axis reactive co-sputtering using planar magnetron sputter guns arranged at 90° intervals around a rectangular substrate. The Zr, Ti and Sn guns were typically run at 150 W, 75 W, and 20 W of radio frequency power, respectively, so as to provide the desired Zr/Ti/Sn composition at the substrate midpoint. Depositions were performed at a total pressure (Ar + O₂) of 30 mT. The substrates were Si wafers coated with 600 Å of TiN. The substrate holder was provided with 10 W of radio frequency power during the deposition.

The electrical properties of the resulting films were evaluated using a scanning Hg-probe instrument. The capacitance of Hg/Zr_xSn_yTi_zO_{2-δ}/TiN capacitors (area $A = 4.05 \times 10^{-3} \text{ cm}^2$) was measured at a frequency of 10 kHz and a signal level of 1 V_{ac}. The breakdown voltage (defined by a current density of $2.5 \times 10^{-4} \text{ A cm}^{-2}$) was obtained by applying a d.c. voltage step ramp and measuring the current with an electrometer. The Hg electrode is biased negative with respect to the TiN electrode, so we injected electrons from the Hg electrode into the Zr_xSn_yTi_zO_{2-δ} film. The capacitance and breakdown voltage were measured at $\sim 4,000$ points in a $66 \times 63 \text{ mm}$ rectangle. In a control experiment, we measured Ta oxide films produced by furnace annealing, which show extremely uniform values of V_{br} and C , demonstrating the reliability of the Hg probe technique. Some Zr_xSn_yTi_zO_{2-δ} films were also evaluated with photolithographically defined Al–0.5% Cu counterelectrodes ranging from $100 \times 100 \mu\text{m}$ to $500 \times 500 \mu\text{m}$ in area, with results consistent with those obtained by the Hg probe technique.

The C and V_{br} data are multiplied point-by-point to obtain the value CV_{br}/A . The FOM is shown in Fig. 1a for a sample deposited at $T_s \sim 200^\circ\text{C}$, and with 40% oxygen in the sputtering ambient (optimum conditions). The properties of a-Zr_xSn_yTi_zO_{2-δ} thin films depend strongly on the deposition conditions. Moreover, the region of composition that yields the highest values for the FOM also depends on the deposition parameters. Thus, full compositional scans for films prepared under various processing conditions proved essential.

The composition was inferred as a function of position using Rutherford backscattering spectroscopy together with independent calibration runs, allowing a mapping of the FOM data onto a conventional ternary phase diagram (Fig. 1b). Because more than one original data point can be mapped onto a single pixel in composition space, we have chosen the highest value in the set. The data from other runs, in which the relative rates of Ti, Zr and Sn were varied by adjusting the radio frequency power applied to the guns, are consistent with Fig. 1b, so we infer that it accurately represents the region of composition space with the highest FOMs. In particular, the region with $z > 0.7$ was determined to have a monotonically decreasing FOM as z increases, which is consistent with the known low E_{br} associated with a-TiO_x. The optimum composition range is shown schematically in Fig. 2.

An I – V characteristic for a film with the approximate composition (Zr_{0.15}Sn_{0.3}Ti_{0.55})₂O_{4-δ} is shown in Fig. 3. The leakage current at the benchmark field 1.5 MV cm^{-2} is $1 \times 10^{-6} \text{ A cm}^{-2}$, slightly above

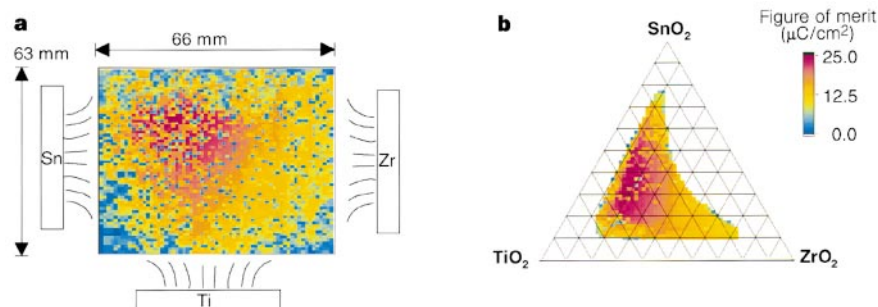


Figure 1 Figure of merit of a-Zr_xSn_yTi_zO_{2-δ} dielectric thin films. **a**, FOM as a function of position on the $6.2 \times 6.5 \text{ cm}$ substrate, obtained by multiplying the capacitance and V_{br} measured at each point, and normalizing to the capacitor area. Note that $CV_{br}/A = \epsilon_r \epsilon_0 E_{br}$. The location of the Zr, Ti and Sn guns is indicated schematically with respect to the position of the substrate. **b**, Same data mapped onto a conventional ternary composition diagram.

the target range for DRAM applications ($1\text{--}5 \times 10^{-7} \text{ A cm}^{-2}$). At this composition, we find $\epsilon_r = 62.5$ and the FOM is $24.3 \mu\text{C cm}^{-2}$.

These results should be contrasted with those of bulk crystalline $\text{Zr}_x\text{Sn}_y\text{Ti}_z\text{O}_4$ ($x + y + z = 2$) (ref. 12). A single homogeneous solid-solution phase, isomorphous with the orthorhombic $\alpha\text{-PbO}_2$ structure, was found near the composition ZrTiO_4 , as indicated in Fig. 2. The remainder of the $\text{Zr}_x\text{Sn}_y\text{Ti}_z\text{O}_4$ phase diagram consists of multiphase mixtures of the ternary phase and the endmembers. Zr–Sn–Ti–O polycrystalline ceramics are well known for their superior microwave properties, including a moderate dielectric constant, very low dissipation factor, and low temperature coefficient of permittivity¹³. The properties of $\text{Zr}_{0.875}\text{Sn}_{0.25}\text{Ti}_{0.875}\text{O}_2$, for example, were studied carefully as a function of doping and sintering. A permittivity of ~ 36 and values for the dissipation factor ($\tan \delta \sim 1/Q$) as low as 7×10^{-5} were obtained. The indicated composition is typical of those used for applications, which are invariably chosen to be within the homogeneous phase field. Note that the optimum compositions in our amorphous films are outside this composition field that yields homogeneous crystalline phases.

Amorphous thin films of $(\text{Zr}, \text{Sn})\text{TiO}_4$ with a particular but unspecified Zr:Sn composition have been studied¹⁴: a value of $\epsilon_r = 27$ was inferred in amorphous films prepared by radio frequency magnetron sputtering at 200°C ; a value of $\epsilon_r = 38$ was found in crystalline films, that is, the bulk value. For approximately the same composition, however, we observe $\epsilon \sim 50$. This is a very substantial difference. At present we cannot explain the discrepancy between our data and those described in ref. 14. We do observe a

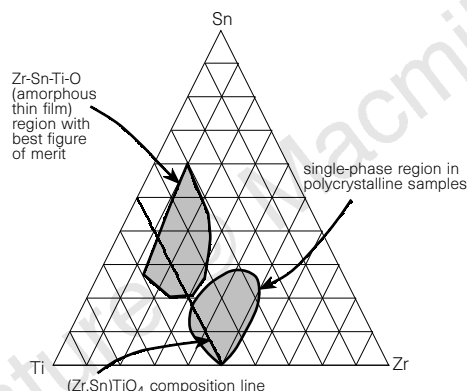


Figure 2 Representation of the ternary phase diagram of Zr–Ti–Sn–O. Shown is the region where single-phase crystalline material is obtained (horizontal hatching) compared to the region where the best thin-film properties are obtained (diagonal hatching), as inferred from Fig. 1b. The dark line indicates the $(\text{Zr}, \text{Sn})\text{TiO}_4$ line composition, which reflects prior art shown in ref. 14.

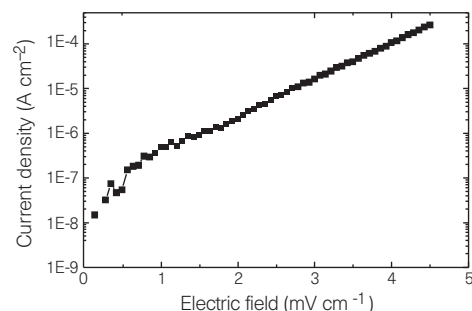


Figure 3 A typical current–voltage characteristic for a film with the approximate composition $\text{Zr}_{0.3}\text{Ti}_{0.9}\text{Sn}_{0.8}\text{O}_2$. The dielectric constant in this film was $\epsilon_r = 62.5$ (using the thickness determined by Rutherford backscattering, RBS, as 70 nm). The FOM for this film was $24.3 \mu\text{C cm}^{-2}$.

decrease in the dielectric constant for deposition at higher temperatures, namely for films with a presumably greater degree of local order, so perhaps the films deposited by Nakagawara *et al.*¹⁴ also have a higher degree of order than our films.

This work demonstrates the value of the CCS technique in the search for metastable materials whose properties are sensitive to preparation conditions. Many issues regarding the development of this material for practical application remain to be addressed. For example, fully optimized uniform films need to be made using a single target with on-axis sputtering (or using CVD), and the properties of films $\sim 10\text{--}20 \text{ nm}$ thick (one fifth the present value) must be evaluated. Also, conditions that lead to a lower leakage current must be found and capacitors made using standard electrodes (for example, TiN) must be subjected to standard post-processing and evaluated for reliability using accelerated testing. Although Zr–Sn–Ti–O films must be further developed, our preliminary results are encouraging. □

Received 22 August; accepted 22 December 1997.

1. *The National Technology Roadmap for Semiconductors* page 123 (Sematech, Austin, 1994); also <http://www.sematech.org/public/roadmap/ntsr94.pdf>.
2. El-Kareh, B., Bronner, G. B. & Schuster, S. E. The evolution of DRAM cell technology. *Solid State Technol.* 89–101 (May 1997).
3. Gerstenberg, D. in *Handbook of Thin Film Technology* (eds Maissel, L. & Glang, R.) Ch. 19, page 8 (McGraw-Hill, New York, 1970).
4. Waser, R., Baiatu, T. & Härdtl, K.-H. DC electrical degradation of perovskite-type titanates: I, ceramics. *J. Am. Ceram. Soc.* 73, 1645–1653 (1990).
5. Baiatu, T., Waser, R. & Härdtl, K.-H. DC electrical degradation of perovskite-type titanates: I, Ceramics. *J. Am. Ceram. Soc.* 73, 1663–1673 (1990).
6. Kahawara, T., Yamamuka, M., Yuuki, A. & Ono, K. (Ba, Sr)TiO₃ films prepared by liquid source chemical vapor deposition on Ru electrodes. *Jpn. J. Appl. Phys.* 35, 4880–4885 (1996).
7. Park, S. O. *et al.* Fabrication and electrical characterization of Pt/(Ba, Sr)TiO₃/Pt capacitors for ultralarge-scale integrated dynamic random access memory applications. *Jpn. J. Appl. Phys.* 35, 1548–1552 (1996).
8. Basceri, C., Streiffer, S. K., Kingon, A. I. & Waser, R. The dielectric response as a function of temperature and film thickness of fiber-textured (Ba, Sr)TiO₃ thin films grown by chemical vapor deposition. *J. Appl. Phys.* 82, 2497–2504 (1997).
9. Xiang, X.-D. *et al.* *Science* 268, 1738–1740 (1995).
10. Sawatzky, E. & Kay, E. Cation deficiencies in RF sputtered gadolinium iron garnet films. *IBM J. Res. Dev.* 13, 696–702 (1969).
11. Hanak, J. J. The ‘multi-sample concept’ in materials research: synthesis, compositional analysis and testing of entire multicomponent systems. *J. Mater. Sci.* 5, 964–971 (1970).
12. Wolfram, G. & Göbel, H. Existence range, structural and dielectric properties of $\text{Zr}_x\text{Ti}_y\text{Sn}_z\text{O}_4$ ceramics ($x + y + z = 2$). *Mater. Res. Bull.* 16, 1455–1463 (1981).
13. Iddles, D. M., Bell, A. J. & Moulson, A. J. Relationships between dopants, microstructure and the microwave dielectric properties of $\text{ZrO}_2\text{--TiO}_2\text{--SnO}_2$ ceramics. *J. Mater. Sci.* 27, 6303–6310 (1992).
14. Nakagawara, O. *et al.* Electrical properties of $(\text{Zr}, \text{Sn})\text{TiO}_4$ dielectric thin film prepared by pulsed laser deposition. *J. Appl. Phys.* 80, 388–392 (1996).

Acknowledgements. We thank H. Huggins for TiN-coated Si substrates; Y.-H. Wong and D. W. Murphy for technical discussions; and R. A. Laudise for enthusiastic support of this project.

Correspondence and requests for materials should be addressed to R.B.v.D.

Decompression-induced melting of ice IV and the liquid–liquid transition in water

Osamu Mishima* & H. Eugene Stanley†

* National Institute for Research in Inorganic Materials (NIRIM), 1-1, Namiki, Tsukuba, Ibaraki 305-0044, Japan

† Center for Polymer Studies and Department of Physics, Boston University, Boston, Massachusetts 02215, USA

Although liquid water has been the focus of intensive research for over 100 years, a coherent physical picture that unifies all of the known anomalies of this liquid^{1–3} is still lacking. Some of these anomalies occur in the supercooled region, and have been rationalized on the grounds of a possible retracing of the liquid–gas spinodal (metastability limit) line into the supercooled liquid region^{4–7} or alternatively the presence of a line of first-order liquid–liquid phase transitions in this region which ends in a critical point^{8–14}. But these ideas remain untested experimentally,

in part because supercooled water can be probed only above the homogeneous nucleation temperature T_H at which water spontaneously crystallizes. Here we report an experimental approach that is not restricted by the barrier imposed by T_H , involving measurement of the decomposition-induced melting curves of several high-pressure phases of ice in small emulsified droplets. We find that the melting curve for ice IV seems to undergo a discontinuity at precisely the location proposed for the line of liquid–liquid phase transitions⁸. This is consistent with, but does not prove, the coexistence of two different phases of (supercooled) liquid water. From the experimental data we calculate a possible Gibbs potential surface and a corresponding equation of state for water, from the forms of which we estimate the coordinates of the liquid–liquid critical point to be at pressure $P_c \approx 0.1$ GPa and temperature $T_c \approx 220$ K.

In the liquid–liquid hypothesis for the anomalies of supercooled water⁸, the liquid can exist in two phases of different densities, here denoted the high-density liquid (HDL) and low-density liquid (LDL) by analogy with the known high-density amorphous (HDA) and low-density amorphous (LDA) phases of ice¹⁵. To explore the region of the phase diagram in which this transition is supposed to occur, we have studied water confined to droplets with diameters of the order of 1–10 μm and exposed to high pressures (Fig. 1a). In such small volumes, nucleation of secondary ice phases following melting in the supercooled regime is kinetically suppressed. This procedure is especially useful in detecting metastable melting curves (Fig. 1b, c) and, as it is not affected by the presence of the line of homogeneous nucleation temperatures $T_H(P)$ (ref. 16) (below which melting is immediately followed by freezing), can be applied to provide a partial test for the hypothesis that there exists a line L_1 of liquid–liquid first-order phase transitions somewhere in the region of the P – T plane below the line $T_H(P)$. Any metastable melting line that crosses L_1 should display discontinuous behaviour at L_1 (such as ice IV shown in Fig. 1d) because the melting behaviour of the ice would be different for the two liquids that coexist at L_1 .

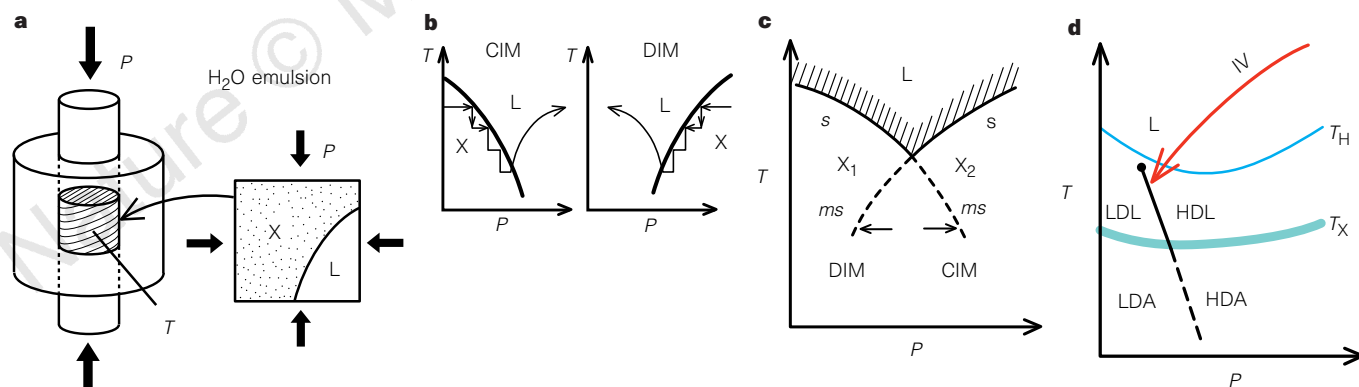


Figure 1 Overall logic of the present experiment. This experiment is designed to probe the region of the temperature–pressure phase diagram—below the nucleation temperature T_H and above the crystallization temperature T_X —where the line L_1 of liquid–liquid phase transitions might exist separating a low-density liquid phase from a high-density liquid phase. **a**, The experimental set-up. The sample, of volume $\sim 1\text{ cm}^3$, consists of 40% v/v of deionized water in a methylcyclohexane/methylcyclopentane mixture and surface stabilizer (sorbitan tristearate) and forms a water emulsion with drop diameter 1–10 μm (ref. 19). Pressure P is applied, keeping the cylinder at a nearly constant low temperature, using an automatic hydraulic pump with a relative error of ~ 2 MPa at 0.1 GPa, and 10 MPa at 1 GPa. The temperature T of the sample is monitored by an alumel–chromel thermocouple with a relative accuracy of ± 0.02 K (ref. 21). The compression (and decompression) rate is held constant at $dP/dt \approx 0.2$ GPa min^{-1} even during transitions. Also shown is a diagram illustrating the case of compression-induced melting from crystal (X) to liquid (L). **b**, The endothermic temperature response of the sample during CIM (compression-induced melting) and DIM (decompression-induced melting). During both CIM and DIM, the crystal X is

We confine 1 cm^3 of water emulsion in an indium capsule and we compress (and decompress) it at a constant rate by a piston cylinder apparatus (Fig. 1a). We measure T using a thermocouple throughout the compression and decompression cycle. As an endothermic (or exothermic) transition results in a change in temperature (Fig. 1b), we can determine with high accuracy the location of the CIM (compression-induced melting) and DIM (decompression-induced melting) lines¹⁶. The present experiments reveal that discontinuous behaviour does occur on melting (Fig. 2). In the region of overlap, the DIM lines agree with the previously reported melting lines^{17,18}, confirming that they are indeed the equilibrium stable melting line and the equilibrium metastable melting line. Moreover, they disappear when we compress an emulsion-carrier sample without water. We confirm the production of the supercooled liquid on further decompression by finding an exothermic transition at the T_H line below 0.2 GPa, which indicates freezing of the liquid¹⁹. Crystalline ice appears, probably due to immediate freezing after melting occurs (below T_H , the transition to another crystalline phase should occur on the DIM line because the liquid produced by DIM should crystallize homogeneously below this temperature).

We extend the T_H line to 1.5 GPa by compressing the supercooled liquid and detecting its freezing on the T_H (Fig. 2a, centre panel). We also observe the CIM line for ice I_h , consistent with previous results¹⁶ (note that we replace the nominal pressure reported in ref. 16 by the real pressure). We find the crystallization line, T_X , of emulsified HDA is higher by ~ 20 K than that of bulk HDA around 1.5 GPa; only the results are shown in Fig. 2a, right panel. The long-dashed line and the horizontal ‘error bar’ around 0.2 GPa in Fig. 2a, right panel, indicate the proposed²⁰ and the experimentally obtained²¹ location of the LDA/HDA ‘equilibrium’ boundary. The long-dashed line around 0.5 GPa indicates the extrapolated I_h /HDA ‘equilibrium’ boundary.

We also find what appear to be two ‘possible new phases’ (PNP), denoted PNP-XIII and PNP-XIV. The DIM line for PNP-XIV is useful because it serves as a ‘control’ for the discontinuous

forced to melt by the pressure, which reduces the sample temperature because the melt absorbs the latent heat quasi-adiabatically. The cooled sample melts by further compression (or decompression). The diagram shows this endothermic temperature response to a sequence of four infinitesimal pressure increments during CIM (and a corresponding temperature response to a sequence of four infinitesimal pressure decrements during DIM). After the entire sample is transformed, the temperature returns to the cylinder temperature. **c**, Typical phase diagram showing two stable melting lines (solid curves, denoted s) separating two stable crystalline phases X_1 and X_2 from the stable liquid (the hatched region). Also shown are two metastable melting lines (dashed curves, denoted ms), one of which can be located using DIM and the other using CIM. Use of an emulsion suppresses the direct X_1/X_2 transitions. **d**, Schematic illustration of the present experiments. We measure the DIM line of ice IV, and search for a kink. We choose ice IV (ref. 29) because its metastable melting line is readily measurable by DIM; it appears to extrapolate to the region of the hypothesized line L_1 of liquid–liquid (LDL/HDL) transitions located between T_H and T_X and terminating at an apparent critical point (the black point).

behaviour shown by the DIM line of ice IV. We obtain PNP-XIII by the decompression-induced transition of the PNP-XIV emulsion annealed around 250 K. We obtain PNP-XIV (1) by heating emulsified HDA above T_X around 1.5 GPa, (2) by compressing super-cooled liquid beyond the T_H line around 230 K, or (3) by cooling the liquid below T_H above ~ 1 GPa.

The smoothness of the PNP-XIV DIM line above T_X implies that L_1 , if it exists, lies in a different region of the P - T diagram traversed by the DIM line of PNP-XIV. PNP-XIV may have a rather ordered structure because of the endothermic nature of the melting and immediate freezing transition below T_H (Fig. 2b, left panel) which indicates that the transition is apparently from a relatively ordered

crystalline phase to a more disordered crystalline phase with net heat absorption.

We note that the DIM line of ice IV shows a sharp kink at 0.1 GPa and 215 K. To interpret this sharp kink, we calculate the Gibbs potential surface for liquid water, $G_L(P, T)$, and we consider if this surface is consistent with the possible existence of L_1 . We shall see that the kink in the DIM line of ice IV may occur precisely at the line L_1 of the hypothesized liquid-liquid transition. As with all phase-transition data, our data are also consistent with the possibility²² of no singularity at all, but rather a sharp but continuous change in the behaviour of the relevant quantities—as there must be experimental error bars on data points (and since the number of data points is

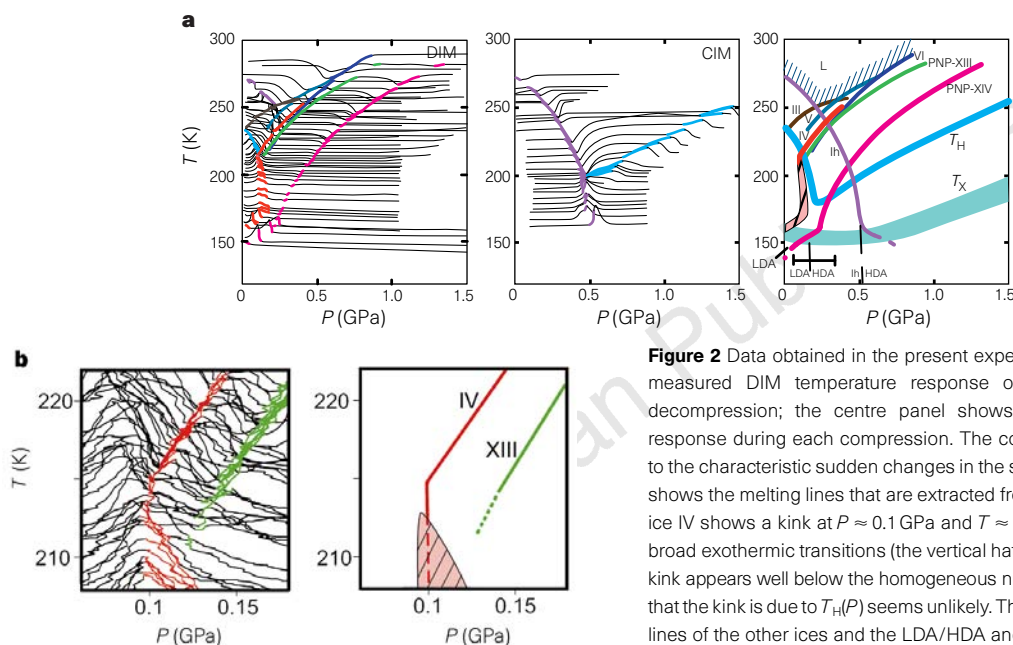


Figure 2 Data obtained in the present experiment. **a**, The left panel shows the measured DIM temperature response of the ice emulsion during each decompression; the centre panel shows the measured CIM temperature response during each compression. The colour-highlighted traces correspond to the characteristic sudden changes in the sample temperature. The right panel shows the melting lines that are extracted from the data shown. The DIM line of ice IV shows a kink at $P \approx 0.1$ GPa and $T \approx 215$ K, and is followed by the line of broad exothermic transitions (the vertical hatched region). We note also that this kink appears well below the homogeneous nucleation line $T_H(P)$, so the possibility that the kink is due to $T_H(P)$ seems unlikely. The right panel also shows the melting lines of the other ices and the LDA/HDA and I_h /HDA transition lines²⁰ (the long-dashed lines). The horizontal bar indicates the range of metastability found for the LDA/HDA transition. **b**, Enlargement of the region of **a** in which the kink in the DIM curve of ice IV occurs (compiling the temperature responses forms the DIM lines of ices).

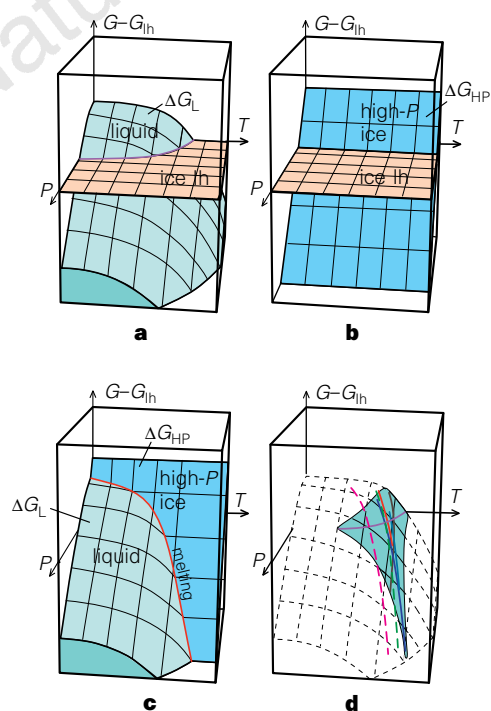


Figure 3 Gibbs potential surface for liquid water probed in the present experiments for both its stable and metastable regions. **a**, Schematic Gibbs potential of liquid water relative to that of ice I_h , $\Delta G_L \equiv G_L - G_{I_h}$. The intersection line of ΔG_L and the basal plane defines the melting line of ice I_h . **b**, Gibbs potential of a typical high-pressure ice relative to that of ice I_h , $\Delta G_{HP} \equiv G_{HP} - G_{I_h}$. **c**, The melting line for the high-pressure ice, defined by the intersection between the Gibbs potential surface of the liquid and that of the high-pressure ice. We obtain numerically the ΔG_{HP} along the melting line which must equal ΔG_L along the melting line. **d**, Schematic construction of the ΔG_L surface (the dark region) by smooth graphical interpolation between the ΔG_L potentials along the melting lines for different ices (the thick solid lines). Each of the nearly vertical thick solid lines is a melting line. The horizontal line is the melting line of ice I_h where the ΔG_L is zero (**a**). These lines locate on a surface (the ΔG_L surface: the dark region). Once we know the ΔG_L surface, we calculate the ΔG_{HP} plane for PNP-XIII and PNP-XIV, following the reverse procedure using the melting lines of PNP ice. We also estimate, and show by the thick dashed lines, the ΔG_L potentials along the melting lines of PNP ices on these ΔG_{HP} planes.

finite, not infinite) there is no *a priori* way to distinguish a function with a sharp discontinuous 'step' from a continuous function with a sharp but still continuous behaviour that merely resembles a step. An example of such a function is $y = \tanh 100x$ which appears to jump discontinuously from -1 for negative x to $+1$ for positive x , yet in fact is a continuous function.

The general procedure we shall follow is based on the fact that along the melting line of any solid ice phase (ice I_h , or any of the high-pressure ice phases), the Gibbs potentials of the coexisting ice and liquid phases must be identical. Hence we can obtain the Gibbs potential of the liquid phase along the melting line of each ice phase by evaluating the Gibbs potential of that ice phase along its melting line.

The Gibbs potential has no natural zero; in this study, we construct $\Delta G_L \equiv G_L - G_{I_h}$, the Gibbs potential of the liquid relative to that of ice I_h (Fig. 3a). Hence the slope of this surface with respect to P must be the difference in specific volumes of the liquid and ice I_h , and the slope with respect to T must be the difference in entropies of the liquid and ice I_h . One advantage of this construction is that the melting line of ice I_h is a curve that is constrained to lie in the plane $\Delta G_L = 0$ (Fig. 3a).

We now consider the Gibbs potential of one of the high-pressure ices relative to that of ice I_h , $\Delta G_{HP} \equiv G_{HP} - G_{I_h}$ (Fig. 3b). We can

obtain the ΔG_{HP} surface by noting that to a good approximation it is a plane, because (1) the difference in specific volume between the high-pressure ice and ice I_h , $V_{HP} - V_{I_h} \equiv [\partial(G_{HP} - G_{I_h})/\partial P]_T$ is roughly constant over the relevant parameter range (0–1.5 GPa and 77–300 K), and (2) the difference in entropy between the high-pressure ice and ice I_h —that is, $S_{HP} - S_{I_h} \equiv -[\partial(G_{HP} - G_{I_h})/\partial T]_P$ —is relatively small²³ (so the pressure at the phase boundary between the high-pressure ice and ice I_h is roughly constant and independent of temperature, and hence the line of intersection with the basal plane is approximately parallel to the T axis). This plane can be determined by knowing any three points or, equivalently, by knowing one point—say, the pressure at the plane's intersection with the basal plane—and the slopes of the plane in the temperature and pressure directions. We therefore calculate the ΔG_{HP} plane from its slope and the pressure at the intersection; we take the slope, $[\partial(\Delta G_{HP})/\partial P]_T = V_{HP} - V_{I_h}$, from ref. 24, and the pressure from that at the intersection between the melting line of the high-pressure ice and that of ice I_h (Fig. 2a, right panel); this is the pressure of the triple point at which three phases (liquid, ice I_h , and the high-pressure ice) coexist.

Figure 3c shows how we calculate ΔG_L along the melting line from the known ΔG_{HP} plane and the known melting line of a given high-pressure ice phase—because $G_L = G_{HP}$ along this line. We can

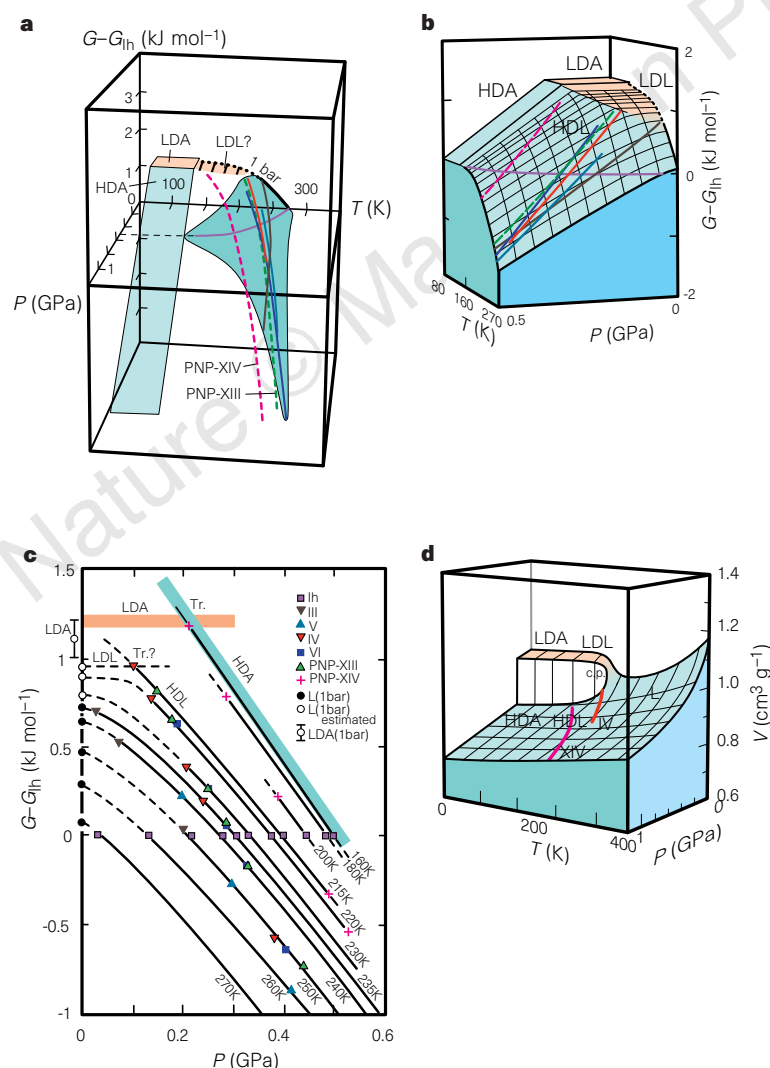


Figure 4 Application of data obtained in the present experiment to reconstruct the Gibbs potential surface and the thermodynamic equation of state $V = V(P, T)$. **a**, Comparison of the ΔG_L surface evaluated by the method of Fig. 3d (the dark region) with (1) the Gibbs potential surfaces of LDA and HDA relative to that of ice I_h , and (2) the ΔG_L at 1 bar. The LDA Gibbs potential at 150 K and 1 bar is known experimentally to be $30 \pm 0.1 \text{ kJ mol}^{-1}$. Also shown are ΔG_L along the melting lines of PNP ices. We find for PNP-XIII a density of 1.25 g cm^{-3} , while for PNP-XIV, we find 1.27 g cm^{-3} . The value of ΔG_L at 1 bar for $T > 235 \text{ K}$ (the thick solid line) we obtain from refs 25, 26, and we assume that this Gibbs potential at 1 bar is connected with the LDA potential³⁰ via the possible low-density liquid (LDL) state—because the liquid above 235 K, cooled extremely rapidly at 1 bar, becomes a glass which resembles LDA^{31,32}. **b**, The $\Delta G_L(P, T)$ surface in the 80–270 K and 0–0.5 GPa region with constant- P and constant- T lines at 50 MPa and 10 K intervals, as evaluated from experimental data. **c**, Dependence of ΔG_L on pressure at some fixed temperatures (also evaluated from experimental data), which may be read in conjunction with **b**: the thin solid constant- T lines in **b** correspond to the lines in this figure with a reverse- P -axis direction. The filled and empty circles at 0 GPa correspond to the measured and estimated potentials at corresponding temperatures at 1 bar. The LDA Gibbs potential at 1 bar is the experimental value³⁰. We can draw smooth lines by interpolation between the Gibbs potential points of the ice phases, and so can construct an accurate ΔG_L surface. To connect the ΔG_L at high pressure with that at 1 bar, a rapid slope change (corresponding to a possible liquid-liquid phase transition, labelled tr) should appear below about 220 K. **d**, Plausible qualitative equation of state $V(P, T)$ of liquid, given by the partial derivative $[\partial(\Delta G_L)/\partial P]_T$ and by adding the specific volume of ice I_h . The specific volumes of the amorphous phases are known for the region below T_X (ref. 21) and for the stable high-temperature liquid³³. Solid lines are the specific volume along the melting lines of ice IV and XIV. The high-temperature liquid appears to separate into two low-temperature liquid phases just below the critical point located at around 0.1 GPa and 220 K; we emphasize that the data cannot locate the coordinates of the critical point, denoted c.p., with high accuracy owing to the possibility that the phase transition line might have a “hook” in it⁹. These two liquid phases are continuous with the two amorphous phases that are known to exist below $\sim 150 \text{ K}$. We note that this phase transition surface differs from that of a typical liquid only in the presence of this critical point—which in turn arises because below the line of density maxima the fluctuations in specific volume and in entropy are anti-correlated by definition.

repeat this procedure for each of the high-pressure ice phases, and thereby obtain values of the Gibbs potential along an entire family of lines (Fig. 3d). Using this family of lines, we then construct the ΔG_L surface (Fig. 3d) by interpolation between the known values of ΔG_L along this family of lines.

Figure 4a shows a portion of the ΔG_L surface. As an accuracy check, we note that this surface includes on it the previously reported^{25,26} line at 1 bar. Also shown are the ΔG surfaces of LDA and HDA. Thus we can calculate the entire ΔG_L surface by connecting these individual lines, and taking the slopes of the LDA phase and the hypothesized LDL phase to be similar (compare the hatched area of Fig. 4a). In this way, we find the LDA and HDA Gibbs potential surfaces, assuming that they are planes and that the I_h -HDA 'equilibrium' transition pressure is 0.5 GPa and the LDA/HDA 'equilibrium' pressure is 0.2 GPa (refs 20, 21; Fig. 2a, right panel). The slope $[\partial(G - G_h)/\partial P]_T$ of LDA and HDA we obtain from their known specific volumes¹⁵. We note that the LDA plane is almost parallel to the basal plane because the specific volume of LDA and that of ice I_h are about the same¹⁵. The thick solid lines in Fig. 4a (except the line at 1 bar) are ΔG_L along the melting lines of high-pressure ices, and the thick broken lines (except the line at 1 bar) are those of the PNP ices. The error in the Gibbs potential is much less than 0.1 kJ mol⁻¹.

Figure 4b shows the results of a detailed calculation of the ΔG_L surface for T between 80 and 270 K and P between 0 and 0.5 GPa. We note that the intersection between the nearly flat LDL surface and the slope of the HDL Gibbs potential makes a distinct 'crease' which starts at the LDA/HDA boundary and continues to a point with coordinates ~ 0.1 GPa and ~ 220 K. This behaviour is consistent with the presence of a first-order transition line, as the specific volume—the first derivative of the Gibbs potential—is discontinuous across the crease.

To further test this possibility, we show in Fig. 4c the pressure dependence of ΔG_L at fixed T . For $T > 235$ K, the ΔG_L at high P extrapolates smoothly to the independently calculated ΔG_L at 1 bar. For the 220 K isotherm, the slope changes rapidly around 0.1 GPa in order to connect to the 1 bar value of ΔG_L , while at 215 K, a crease in the surface must occur if the values of ΔG_L at 1 bar and 0.1 GPa are to be on the same curve.

We note that this point (0.1 GPa, 215 K) is precisely the point where the melting curve of ice IV undergoes a sharp discontinuity, so $\Delta G_L = \Delta G_{IV}$. Although this could be a coincidence, it could also arise because of the presence of a line of first-order liquid-liquid phase transitions.

In Fig. 4d we show a possible equation of state $V = V(P, T)$ that is consistent with experimental data. The specific volumes of the two liquid phases LDL and HDL appear to be continuous with the specific volumes of LDA and HDA. Further, the anomalous behaviour of supercooled D₂O water in terms of both liquid structure²⁷ and molecular relaxation²⁸ are also consistent with the existence of a critical point with approximately the same coordinates. □

Received 6 January; accepted 9 February 1998.

1. Debenedetti, P. G. *Metastable Liquids* (Princeton Univ. Press, 1997).
2. Angell, C. A. Formation of glasses from liquids and biopolymers. *Science* **267**, 1924–1935 (1995).
3. Fourkas, J. T., Kivelson, D., Mohanty, U. & Nelson, K. A. (eds) *Supercooled Liquids: Advances and Novel Applications* (ACS Books, Washington DC, 1997).
4. Speedy, R. J. & Angell, C. A. Isothermal compressibility of supercooled water and evidence for a thermodynamic singularity at -45°C . *J. Chem. Phys.* **65**, 851–858 (1976).
5. Speedy, R. J. Stability-limit conjecture. An interpretation of the properties of water. *J. Phys. Chem.* **86**, 982–991 (1982).
6. Borick, S. S. & Debenedetti, P. G. Equilibrium, stability and density anomalies in a lattice model with core softening and directional bonding. *J. Phys. Chem.* **97**, 6292–6303 (1993).
7. Borick, S. S., Debenedetti, P. G. & Sastry, S. A lattice model of network-forming fluids with orientation-dependent bonding: equilibrium, stability, and implications from the phase behavior of supercooled water. *J. Phys. Chem.* **99**, 3781–3793 (1995).
8. Poole, P. H., Sciortino, F., Essmann, U. & Stanley, H. E. Phase behaviour of metastable water. *Nature* **360**, 324–328 (1992).
9. Tanaka, H. A self-consistent phase diagram for supercooled water. *Nature* **380**, 328–331 (1996).
10. Tanaka, H. Phase behaviors of supercooled water: reconciling a critical point of amorphous ices with spinodal instability. *J. Chem. Phys.* **105**, 5099–5111 (1996).
11. Roberts, C. J., Panagiotopoulos, A. Z. & Debenedetti, P. G. Liquid-liquid immiscibility in pure fluids: polymorphism in simulations of a network-forming fluid. *Phys. Rev. Lett.* **77**, 4386–4389 (1996).

12. Roberts, C. J. & Debenedetti, P. G. Polyamorphism and density anomalies in network-forming fluids: zeroth- and first-order approximations. *J. Chem. Phys.* **105**, 658–672 (1996).
13. Harrington, S., Zhang, R., Poole, P. H., Sciortino, F. & Stanley, H. E. Liquid-liquid phase transition: Evidence from simulations. *Phys. Rev. Lett.* **78**, 2409–2412 (1997).
14. Sciortino, F., Poole, P. H., Essmann, U. & Stanley, H. E. Line of compressibility maxima in the phase diagram of supercooled water. *Phys. Rev. E* **55**, 727–737 (1997).
15. Mishima, O., Calvert, L. D. & Whalley, E. An apparently first-order transition between two amorphous phases of ice induced by pressure. *Nature* **314**, 76–78 (1985).
16. Mishima, O. Relationship between melting and amorphization of ice. *Nature* **384**, 546–549 (1996).
17. Bridgman, P. W. Water, in the liquid and five solid forms, under pressure. *Proc. Am. Acad. Arts Sci.* **47**, 441–558 (1911).
18. Evans, L. F. Selective nucleation of the high pressure ices. *J. Appl. Phys.* **38**, 4930–4932 (1967).
19. Kanno, H., Speedy, R. & Angell, C. A. Supercooling of water to -92°C under pressure. *Science* **189**, 880–881 (1975).
20. Whalley, E., Klug, D. D. & Handa, Y. P. Entropy of amorphous ice. *Nature* **342**, 782–783 (1989).
21. Mishima, O. Reversible first-order transition between two water amorphs at ~ 0.2 GPa and ~ 135 K. *J. Chem. Phys.* **100**, 5910–5912 (1994).
22. Sastry, S., Debenedetti, P., Sciortino, F. & Stanley, H. E. Singularity-free interpretation of the thermodynamics of supercooled water. *Phys. Rev. E* **53**, 6144–6154 (1996).
23. Fletcher, N. H. *The Chemical Physics of Ice* (Cambridge Univ. Press, 1970).
24. Kamb, B. Structure of ice VI. *Science* **150**, 205–209 (1965).
25. Speedy, R. J. Thermodynamic properties of supercooled water at 1 atm. *J. Phys. Chem.* **91**, 3354–3358 (1987).
26. Johari, G. P., Fleissner, G., Hallbrucker, A. & Mayer, E. Thermodynamic continuity between glassy and normal water. *J. Phys. Chem.* **98**, 4719–4725 (1994).
27. Bellissent-Funel, M. C. & Bosio, L. A neutron scattering study of liquid D₂O. *J. Chem. Phys.* **102**, 3727–3735 (1995).
28. Lang, E. & Ludemann, H.-D. Pressure and temperature dependence of the longitudinal deuterium relaxation times in supercooled heavy water to 300 MPa and 188 K. *Ber. Bunsenges. Phys. Chem.* **84**, 462–470 (1980).
29. Bridgman, P. W. The pressure-volume-temperature relations of the liquid, and the phase diagram of heavy water. *J. Chem. Phys.* **3**, 597–605 (1935).
30. Speedy, R. J., Debenedetti, P. G., Smith, R. S., Huang, C. & Kay, B. D. The evaporation rate, free energy, and entropy of amorphous water at 150 K. *J. Chem. Phys.* **105**, 240–244 (1996).
31. Bruggeller, P. & Mayer, E. Complete vitrification in pure liquid water and dilute aqueous solutions. *Nature* **288**, 569–571 (1980).
32. Bellissent-Funel, M. C., Bosio, L., Hallbrucker, A., Mayer, E. & Sridi-Dorbez, R. X-ray and neutron scattering studies of the structure of hyperquenched glassy water. *J. Chem. Phys.* **97**, 1282–1286 (1992).
33. Haar, L., Gallagher, J. S. & Kell, G. S. *NBS/NRC Steam Tables* (Hemisphere, Washington DC, 1984).

Acknowledgements. We thank C. A. Angell, K. Aoki, M.-C. Bellissent-Funel, M. Canpolat, H.-D. Ludemann, P. H. Poole, R. Sadri-Lahijany, S. Sastry, F. Sciortino, F. W. Starr and Y. Suzuki for discussions and reading of manuscript drafts. We also thank C. A. Angell for pointing out subtle points that we initially glossed over. This work was supported by CREST (Core Research for Evolutional Science and Technology) of Japan Science and Technology Corporation (JST), BP and the US NSF.

Correspondence and requests for materials should be addressed to O.M. (e-mail: mishima@nirim.go.jp).

Direct measurement of tropospheric ozone distributions from space

Rosemary Munro*, Richard Siddans, William J. Reburn & Brian J. Kerridge

Rutherford Appleton Laboratory, Chilton, Didcot, Oxfordshire, OX11 0QX, UK

The role of ozone in absorbing ultraviolet solar radiation is well known. Ozone also makes a significant contribution to the radiative balance of the upper troposphere and lower stratosphere, such that changes in the distribution of ozone in these atmospheric regions will affect the radiative forcing of climate^{1,2}. Furthermore, tropospheric ozone is the source of the hydroxyl radical which controls the abundance and distribution of many atmospheric constituents, including greenhouse gases such as methane and hydrochlorofluorocarbons. Tropospheric ozone is produced photochemically *in situ* and is also transported down from the stratosphere, but the relative importance of these two sources to its global budget is poorly understood. High-quality tropospheric and lower-stratospheric ozone profile measurements are available from sondes and lidar techniques, but their geographical sampling is very limited. Complementary satellite measurements of the global ozone distribution in this height region are therefore required to quantify ozone's tropospheric budget and its participation in climate-forcing and tropospheric

* Present address: ECMWF, Shinfield Park, Reading, Berkshire, RG2 9AX, UK.

chemistry. Here we present direct measurements of tropospheric ozone concentrations from space, made by the European Space Agency's Global Ozone Monitoring Experiment. These results demonstrate the potential of satellite measurements to provide self-consistent tropospheric and stratospheric ozone distributions on a global scale.

Measurements by the satellite-borne Total Ozone Mapping Spectrometer (TOMS) and Solar Backscatter Ultra Violet (SBUV) instruments have provided ozone total atmospheric column and stratospheric profile information on a global basis for nearly two decades^{3–5}. Ozone in the tropospheric part of the column has previously been inferred indirectly from TOMS and the Stratospheric Aerosol and Gas Experiment (SAGE) observations⁶ and from TOMS observations at different locations⁷ but, before the launch of the Global Ozone Monitoring Experiment (GOME) on the second European Remote-sensing Satellite (ERS-2) in April 1995, individual profiles spanning the troposphere and stratosphere had never been measured from space.

GOME is an ultraviolet/visible spectrometer which measures solar radiation backscattered from the Earth's atmosphere in four contiguous wavelength bands between 237 and 790 nm at moderate resolution (0.2–0.4 nm)⁸. Measurements in bands 1A (237–307 nm) and 2B (312–405 nm) were used in this study, exploiting the ozone Hartley band (240–300 nm) and Huggins bands (310–340 nm), respectively. In the Hartley band, strong wavelength-dependent absorption by ozone determines the altitude from which radiation is backscattered. The resulting wavelength dependence in effective scattering altitude provides profile information down to the ozone peak concentration (as demonstrated by SBUV), but only limited information at lower altitudes. The ability to resolve temperature-dependent spectral structures in the Huggins bands is an added advantage for GOME. With knowledge of the temperature profile, this structure yields additional profile information below the ozone peak (that is, in the troposphere and lower stratosphere): an important advance over SBUV.

In addition to ozone's optical properties, integration times for individual measurements also differ substantially between bands 1A (12 s) and 2B (0.37 s). Hence, each band 2B 'scene' covers only a small fraction of that over which band 1A integrates. For the swath-width used on the days in question, integrations were over scenes of dimensions 960 km (across-track) by 80–120 km (along-track) for band 1A, and 80 km (across-track) by 40 km (along-track) for band 2B. Because of these important differences, a two-step approach was adopted for bands 1A and 2B. Only results from band 2B retrievals, the second step, are described here.

We present here results from the analysis of five band 2B spectra: 4 April 1996 (53° N, 2.5° E) near De Bilt, Holland; 21 October 1995 (53° N, 2° W) near Aberystwyth, UK and 12 November 1995 (52° N,

1.5° E) also near Aberystwyth, UK; 11 January 1996 near Lilongwe, Malawi (14° N, 30.6° E) and 4 July 1996 also near Lilongwe, Malawi (14° N, 33.5° E). Two of those chosen (4 April 1996 and 21 October 1995) have been demonstrated to be cloud-free using 1 × 1 km resolution images from the second Along-Track Scanning Radiometer (ATSR-2), also on board ERS-2⁹. ATSR-2 data indicate that the other Aberystwyth scene (12 November 1995) has 100% low cloud cover. In these three cases, co-located ozone and temperature profiles were available from sondes launched on the same day; sonde temperature profiles above 33 km were blended with those from the UK Meteorological Office analysis for the closest location and time.

The principal input to the ozone retrieval scheme is the reflectance spectrum measured by GOME—the ratio of backscattered (radiance) to direct-Sun (irradiance) spectra which have been calibrated both radiometrically and in wavelength (referred to as the GOME Level-1 Data Product). However, before generating reflectance spectra, it is necessary to first use the direct-Sun spectrum to improve absolute wavelength calibration and knowledge of the spectrometer slit-width. This is achieved by fitting segments of the measured spectrum in a piecewise fashion to calculations which used the Harvard-Smithsonian 0.01-nm-resolution solar reference spectrum (K. Chance, personal communication).

To synthesize reflectance spectra, the GOMETRAN radiative transfer model¹⁰ has been interfaced to the retrieval scheme. This model has a wavelength resolution appropriate to GOME, and incorporates both gaseous absorption and also surface reflection and scattering by air molecules, aerosol and cloud. Multiple scattering is accounted for by a finite-difference method. In the cases presented here a fixed, single profile was adopted for aerosol scattering and cloud has been neglected.

A limited sub-set of wavelengths (between 325.5 and 334.5 nm) has been selected for ozone retrievals from band 2B. To maximize sensitivity to the ozone spectral signature and to minimize sensitivity to scattering and reflecting media, the natural logarithm of reflectance is used as the measured quantity and a second-order polynomial is subtracted from measured and calculated spectra before fitting.

Unwanted spectral structure is present in measured spectra due to NO₂ absorption, the Ring Effect¹⁷, wavelength shifts between measurements of backscattered and direct-Sun spectra, and wavelength offsets with respect to the reference ozone absorption cross-section spectrum. To reduce r.m.s. residuals between measured and calculated spectra to the level needed (<0.3%) to extract useful ozone information from the troposphere, it is necessary to account for NO₂ using a single, fixed profile and to retrieve the other quantities jointly with ozone and surface reflectance.

The retrieval method used is nonlinear optimal estimation^{11,12}. Because retrieval of height-resolved ozone information from an

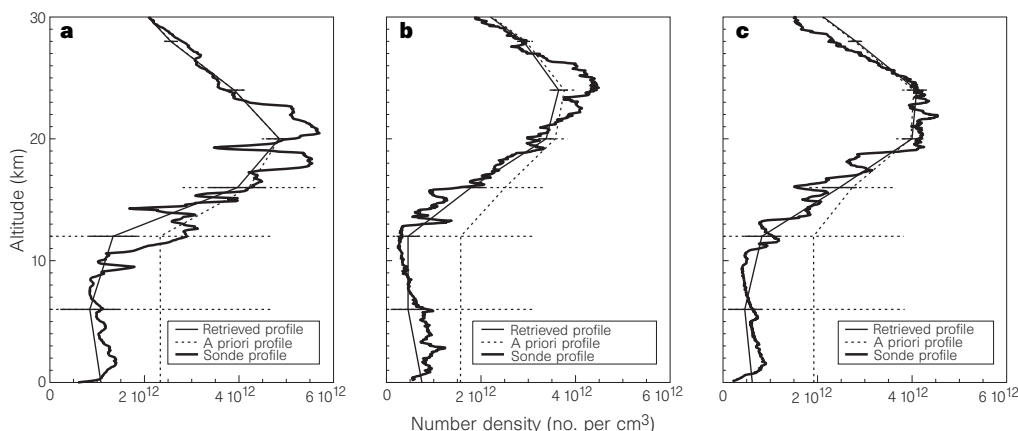


Figure 1 Ozone profiles retrieved from GOME band 2 data, coincident sonde profiles and a priori profiles used as a constraint in the retrieval process. Profiles are shown for two locations and three times: **a**, De Bilt, 4 April 1996; **b**, Aberystwyth, 21 October 1995; and **c**, Aberystwyth, 12 November 1995.

ultraviolet spectrum is a nonlinear underconstrained problem, it is necessary to provide an additional *a priori* constraint (in this case an ozone profile with associated errors), linearize and iterate until convergence is obtained. The solution from such a retrieval combines information from the measurements and the *a priori* profile to produce a solution which is an optimal combination of the two.

At 16 km and higher altitudes, the *a priori* ozone concentration and its uncertainty are those retrieved from the associated band 1A spectrum, thereby constraining the band 2B retrieval in this height range (to $\pm 10\%$ at 20 km and above, and to $\pm 30\%$ at 16 km). To demonstrate clearly the potential of band 2B for retrieval of tropospheric ozone in the examples shown here, the *a priori* concentra-

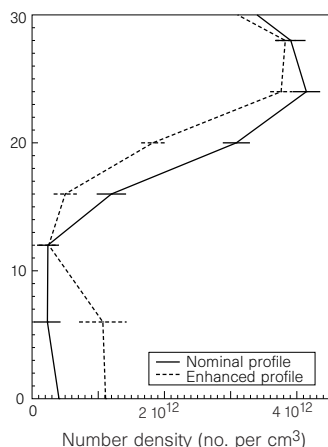


Figure 2 Ozone profiles retrieved from GOME band 2 data. Profiles are for 14° N, 30.6° E, on 4 July 1996 during the summer season and for 14° N, 33.5° E on 11 January 1996 during the winter season, clearly showing detection of enhanced tropospheric ozone levels.

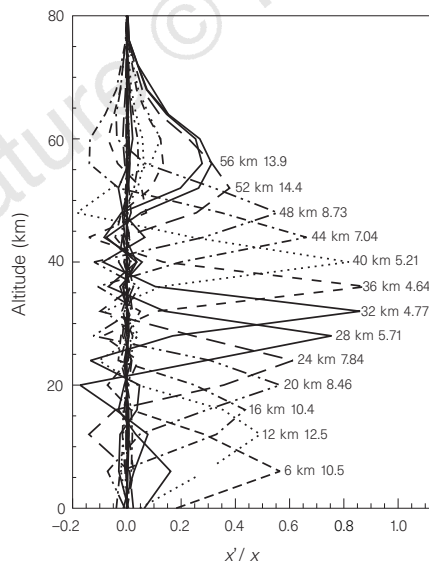


Figure 3 Ozone averaging kernels (vertical resolution functions) for the composite of GOME band 1A and 2B measurements. These show the sensitivity of ozone concentrations retrieved at particular altitudes (x') to perturbations in ozone concentration (x) at all altitudes (that is, $\delta x'/\delta x$). An ideal averaging kernel would be a delta function at the altitude of the associated retrieval level. The extent to which independent information is available about ozone in different atmospheric layers is indicated by the overlap between averaging kernels. The retrieval altitude is shown next to the peak of each kernel, together with its FWHM, which provides a measure of vertical resolution.

tion at 0, 6 and 12 km was specified arbitrarily to be a constant value, and the *a priori* uncertainty was specified to be 100%. Thus the band 2B retrieval was only weakly constrained in the troposphere.

Retrieved and *a priori* ozone profiles for the de Bilt and Aberystwyth scenes are shown in Fig. 1, together with their respective uncertainties and the profiles from co-located ozone sondes launched on the same days. In all three cases, the retrieved profile below 16 km deviates substantially from the *a priori* profile but agrees well with the retrieved sonde profile. It is important to note that, even in the presence of uniform low cloud (Fig. 1c), it has been possible to retrieve reliable tropospheric information.

These three comparisons with sonde profiles demonstrate that, in cloud-free scenes and above low cloud, ozone profiles extending down through the troposphere can be retrieved from GOME measurements; a significant innovation for satellite remote-sensing. If high-level cloud is substantial and unaccounted for, it causes an ozone underestimate by comparison to sondes. However, a technique to identify upper-level cloud has been implemented using GOME spectra. The Ring Effect scaling factor (referred to previously) is proportional to the number of scattering molecules in the optical path which, in the absence of cloud, varies smoothly with latitude. In the presence of upper-level cloud, the number of scattering molecules in the optical path is reduced, so the scaling factor deviates below the expected value.

The geographical and temporal variability of tropospheric ozone is considerable—for example, enhancement of $\sim 200\%$ can occur in biomass-burning regions¹⁵. An example of GOME's ability to detect tropospheric ozone produced from biomass burning over Africa is presented in Fig. 2, where the retrieved profile for 14° N, 30.6° E, on 4 July 1996 during the summer season when biomass burning would be expected to occur ('enhanced profile'), is compared to that for a similar location 14° N, 33.5° E on 11 January 1996 during the winter season ('nominal profile').

An indication of GOME's intrinsic vertical resolution is provided by the full-width at half-maximum (FWHM) of the retrieval averaging kernels¹², examples of which are presented in Fig. 3. Averaging kernels depend on the shape of the ozone profile itself and therefore differ from one profile to the next and between *a priori* and retrieved profiles. Those in Fig. 3 were calculated for a nominal profile. Except at the surface, peak values are >0.5 , demonstrating that, for the assumed *a priori* error of 100%, the retrieved profile at levels above the surface is influenced more strongly by the GOME measurements than by the *a priori* profile. Although these composite averaging kernels for bands 1A and 2B are narrowest (<5 km FWHM) between 30 and 40 km, those associated with tropospheric retrieval levels at 6 and 12 km have well-defined peaks, confirming that information on tropospheric ozone contained in GOME spectra is separable from that on stratospheric ozone. These results are in accord with those from theoretical studies^{14–16} carried out before the launch of GOME.

On a given day, GOME acquires a total of $>28,000$ spectra from ~ 14 orbits, sampling along track every 40 km. The error bars in Figs 1 and 2 indicate the precision obtained for individual profile retrievals from individual GOME spectra, and are proportional to the r.m.s. residuals between measured and calculated spectra. The spectral structure in the residuals is highly reproducible and the r.m.s. values are more than a factor of two higher than the photon noise limit. Once the systematic component has been removed and the r.m.s. reduced, the precision on individual profiles will be increased substantially.

GOME has the potential to deliver unique information on the height-resolved distribution of ozone; this information can be used to investigate the tropospheric budget of ozone and its roles in climate and tropospheric chemistry, in addition to monitoring its stratospheric abundance. To fulfil this potential, a scheme to produce height-resolved data on a large scale is currently under development. This scheme will also exploit co-located high-resolution

(1 × 1 km) images of visible reflectance and infrared brightness temperature from a companion instrument ATSR-2, to estimate cloud fraction, cloud-top height and other cloud and surface parameters. □

Received 29 November 1996; accepted 9 December 1997.

- Houghton, J. T. *et al.* (eds) *Climate Change 1994, Part 1, Radiative Forcing of Climate Change* (Cambridge Univ. Press, 1995).
- Houghton, J. T. *et al.* (eds) *Climate Change 1995* (Cambridge Univ. Press, 1996).
- Hollandsworth, S. M. *et al.* Ozone trends deduced from combined NIMBUS-7 SBUV and NOAA-11 SBUV/2 data. *Geophys. Res. Lett.* **22**, 905–908 (1995).
- McPeters, R. D., Hollandsworth, S. M., Flynn, L. E., Herman, J. R. & Seftor, C. J. Long-term ozone trends derived for the 16-year combined Nimbus 7 Meteor 3 TOMS Version 7 record. *Geophys. Res. Lett.* **23**, 3699–3702 (1996).
- WMO, *Report of the International Ozone Trends Panel 1988* Vol. 1, 17–52 (Rep. No. 18, Global Ozone Research and Monitoring Project, World Meteorological Organization, Geneva, 1988).
- Fishman, J., Watson, C. E., Larsen, J. C. & Logan, J. A. Distribution of tropospheric ozone determined from satellite data. *J. Geophys. Res.* **95**(D4), 3599–3617 (1990).
- Jiang, Y. B. & Yung, Y. L. Concentration of tropospheric ozone from 1979 to 1992 over tropical Pacific South-America from TOMS data. *Science* **272**, 714–716 (1996).
- Hahne, A., Lefebvre, A., Callies, J. & Christensen, B. GOME—The development of a new instrument. *ESA Bull.* **83**, 41–46 (1995).
- Francis, C. R. *et al.* The ERS-2 spacecraft and its payload. *ESA Bull.* **83**, 13–31 (1995).
- Rosnov, V., Diebel, D., Spurr, R. J. D. & Burrows, J. P. GOMETRAN: A radiative transfer model for the satellite project GOME—the plane-parallel version. *J. Geophys. Res.* (in the press).
- Rodgers, C. D. Retrieval of atmospheric temperature and composition from remote measurements of thermal radiation. *Rev. Geophys. Space Phys.* **14**, 609–624 (1976).
- Rodgers, C. D. Characterization and error analysis of profiles retrieved from remote sounding measurements. *J. Geophys. Res.* **95**(D5), 5587–5595 (1990).
- Collins, W. J., Stevens, D. S., Johnson, C. E. & Derwent, R. G. Tropospheric ozone in a global-scale three-dimensional Lagrangian Model and its response to NO_x emission controls. *J. Atmos. Chem.* (in the press).
- Burrows, J. *et al.* *A Study of Methods for the Retrieval of Atmospheric Constituents* (Final Rep. ESA Contract 9687/91/NL/BI, European Space Agency, Noordwijk, The Netherlands, 1994).
- Diebel, D. *et al.* *Detailed Analysis of the Retrieval Algorithms Selected for Level 1-2 Processing of GOME Data* (Final Rep. ESA Contract 10728/94/NL/CN, European Space Agency, Noordwijk, The Netherlands, 1995).
- Chance, K. V., Burrows, J. P., Perner, D. & Schneider, W. Satellite measurements of atmospheric ozone profiles, including tropospheric ozone, from ultraviolet/visible measurements in the nadir geometry: a potential method to retrieve tropospheric ozone. *J. Quant. Spectrosc. Radiat. Transfer* **57**(4), 467–476 (1997).
- Grainger, J. F. & Ring, J. Anomalous Fraunhofer line profiles. *Nature* **193**, 762 (1962).

Acknowledgements. Permission to use GOME data before its public release was granted by ESA, and GOME Level-1 Products were provided by DLR. The GOMETRAN radiative transfer model and the high-resolution solar reference spectrum used in this study were supplied by J. Burrows (Univ. Bremen) and K. Chance (Harvard-Smithsonian Astrophysical Observatory), respectively.

Correspondence and requests for materials should be addressed to R.M. (e-mail: r.munro@ecmwf.int).

Evidence for a late Triassic multiple impact event on Earth

John G. Spray*, Simon P. Kelley† & David B. Rowley‡

* Department of Geology, University of New Brunswick, Fredericton E3B 5A3, Canada

† Department of Earth Sciences, The Open University, Milton Keynes MK7 6AA, UK

‡ Department of Geophysical Sciences, University of Chicago, Chicago, Illinois 60637, USA

Evidence for the collision of fragmented comets or asteroids with some of the larger (jovian) planets and their moons is now well established following the dramatic impact of the disrupted comet Shoemaker–Levy 9 with Jupiter in 1994 (ref. 1). Collisions by fragmented objects result in multiple impacts that can lead to the formation of linear crater chains, or catenae, on planetary surfaces². Here we present evidence for a multiple impact event that occurred on Earth. Five terrestrial impact structures have been found to possess comparable ages (~214 Myr), coincident with the Norian stage of the Triassic period. These craters are Rochechouart (France), Manicouagan and Saint Martin (Canada), Obolon' (Ukraine) and Red Wing (USA). When these impact structures are plotted on a tectonic reconstruction of the North American and Eurasian plates for 214 Myr before present, the three largest structures (Rochechouart, Manicouagan and Saint Martin) are co-latitudinal at 22.8° (within 1.2°, ~110 km), and

span 43.5° of palaeolongitude. These structures may thus represent the remains of a crater chain at least 4,462 km long. The Obolon' and Red Wing craters, on the other hand, lie on great circles of identical declination with Rochechouart and Saint Martin, respectively. We therefore suggest that the five impact structures were formed at the same time (within hours) during a multiple impact event caused by a fragmented comet or asteroid colliding with Earth.

Approximately 150 impact structures are now known on Earth, most of which are <200 Myr old³. They represent a small portion of what would have been a much larger number, most structures having been buried or destroyed via the tectonic and erosional activities of our dynamic planet. The discoveries of an iridium-enriched clay layer⁴ and the 65-Myr-old Chicxulub impact structure in Yucatán, Mexico⁵, both coincident with the Cretaceous/Tertiary (K/T) boundary, have drawn attention to the global environmental damage such impacts can cause and, hence, their association with mass extinction⁶. Geologists have tended to seek single giant impact structures as potential sources of catastrophe on Earth, but the temporal and spatial association of five impact structures described here indicates that the effects of synchronous multiple impact events produced by fragmented comets or asteroids should not be discounted.

The Rochechouart impact structure in France occurs within Hercynian target rocks of the Massif Central. The structure is partly eroded and no impact-related topography remains. A thin (<10 m) impact melt sheet is overlain by a fallback sequence 20–50 m thick, and underlain by allochthonous and autochthonous basal breccias. A shock zoning study and the distribution of various impact lithologies indicate that the structure is ~25 km in diameter⁷. Previous radiometric work has yielded ages of 154–173 Myr (K–Ar)⁸ and 186 Myr (Rb–Sr)⁹, but the recent ⁴⁰Ar/³⁹Ar laser spot fusion dating of pseudotachylyte generated in impact-related basement faults has yielded¹⁰ 214 ± 8 Myr (2σ). This revised age is in keeping with the regional geological setting of the structure, whereas the younger age determinations, which implied a Jurassic origin, are now considered to indicate the timing of subsequent Ar loss and Rb and Sr mobilization. The Manicouagan impact structure in eastern Canada is ~100 km in diameter. The target consists of amphibolite to granulite facies metamorphic rocks and anorthosites of Grenville age (~1 Gyr) with overlying Ordovician carbonates. The anorthositic rocks have been locally shocked to yield diaplectic plagioclase glass (maskelynite)¹¹. High precision U–Pb

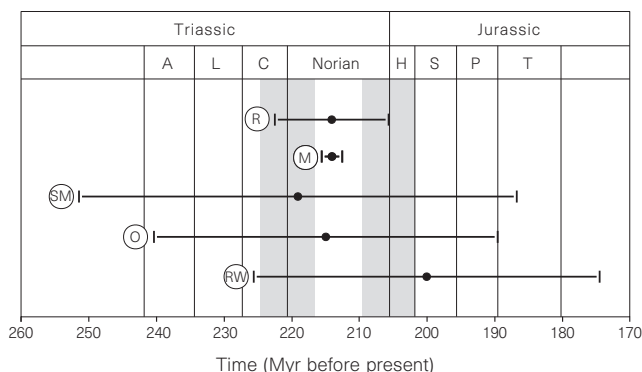


Figure 1 Radiometric and biostratigraphic age data for the five impact structures plotted for the late Triassic to early Jurassic using the timescale of Gradstein *et al.*²². Abbreviations for stages as follows: A, Anisian; L, Ladinian; C, Carnian; H, Hettangian; S, Sinemurian; P, Pliensbachian; T, Toarcian. (Rhaetian substage included in the Norian.) Shaded columns show errors for the Carnian/Norian (±4.4 Myr) and Norian/Hettangian (±4 Myr) boundaries²². Letter abbreviations for the impact structures (given in circles) as follows: R, Rochechouart; M, Manicouagan; SM, Saint Martin; O, Obolon'; RW, Red Wing.

dating of zircons from the >200-m-thick impact melt sheet yields¹² an age of 214 ± 1 Myr (2σ). The Saint Martin impact structure in western Canada is not well exposed, but gravity and magnetic surveys, coupled with the results of a drilling programme in the late 1960s, have revealed its essential structure. It is ~40 km in diameter¹³. The target rocks comprise Superior Province (Archaean) granitic gneisses and overlying Palaeozoic sedimentary cover. The youngest pre-impact material is Devonian, while Jurassic red beds and evaporites constitute the oldest post-impact cover. Rb/Sr dating of the impact melt sheet yields¹⁴ an age of 219 ± 32 Myr (2σ). The Obolon' impact structure of the Ukraine is ~15 km in diameter and buried by 150–300 m of post-impact Middle Jurassic to Quaternary sedimentary rocks¹⁵. Geophysical and bore-hole data indicate that the target lithologies are early Proterozoic migmatitic gneisses and granitoids, overlain by a mid-Carboniferous to Lower Triassic sedimentary cover. Shatter cones, melt-bonded breccias, diaplectic glass and the presence of coesite¹⁶, microdiamonds¹⁷ and planar deformation features in quartz¹⁵ have been identified from drill core. No radiometric constraints are available for Obolon', but stratigraphic evidence supports a formation age¹⁸ of 215 ± 25 Myr. The Red Wing impact structure of the western USA is ~9 km in diameter and buried by ~1.5 km of Jurassic to Neogene post-impact sediments. The impact structure is formed in Silurian to Triassic carbonates, minor sandstones and evaporites, and provides the structural trap to one of the most prolific oil fields in the United States¹⁹. It has a stratigraphically constrained age²⁰ of 200 ± 25 Myr.

Each of the five structures shows features characteristic of hypervelocity impact, including most of the following: impact-generated melt sheets of crustal bulk composition, shatter cones, planar deformation features in minerals, high-pressure polymorphs and diaplectic glasses. All are complex craters and are probably central peak basins²¹, though Manicouagan may be large enough to be a peak-ring or multi-ring basin. Based on the timescale of Gradstein *et al.*²², four of the five impact structures have ages that fall within the Norian stage of the late Triassic, while Red Wing overlaps with the Norian within error (Fig. 1). However, given the possibility of the impacts coinciding to within a few hours, temporal constraints provided by radiometric dating alone cannot prove the synchronicity of the impacts due to experimental uncertainties. The critical test for synchronous impact remains a spatial one.

Using the two best-dated structures as an age constraint (Rochechouart and Manicouagan), the five sites have been plotted on a map showing the reconstructed positions of the North American and Eurasian plates 214 Myr ago²³ (Fig. 2). The three largest impact structures (Rochechouart, Manicouagan and Saint Martin) plot as co-latitudinal at a mean palaeolatitude of 22.8° , and a latitudinal

width of $\sim 1.2^\circ$. This is a remarkably good fit to a small-circle path about the Earth's spin axis. The spread in palaeolongitude is 43.5° (4,462 km). The three aligned structures thus form a crater chain, or catena, like the catenae described on Jupiter's moons². The probability of the three co-latitudinal structures representing a random event was assessed using Monte Carlo simulation. The probability ranges in a nonlinear fashion from $P \approx 0.0003$ to $P \approx 0.083$ for $N = 3$ to $N = 13$, where N is the total number of coeval impactors. These probabilities decrease by a factor of about 10 if the longitudinal range of the co-latitudinal impacts is restricted to $\leq 60^\circ$, which is the case here. It is thus highly unlikely that these represent a random array. The two smallest impact structures, Obolon' (15 km) and Red Wing (9 km), have essentially identical trajectories with respect to the latitude-parallel trajectory of the other three. Obolon' and Rochechouart (the easternmost pair) define a great circle that has a declination of 37.5° , while Red Wing and Saint Martin (the westernmost pair) define a great circle that has a declination of 43.4° . Thus, they have the same sense and essentially the same magnitude of rotation with respect to the small-circle trajectory. If the longitudinal offset of 43.5° is removed for Red Wing and Saint Martin, while maintaining their latitudes, a best-fit great circle with a declination of 37.2° is obtained for the four 'end' craters (Red Wing, Saint Martin, Rochechouart and Obolon') (Fig. 2). Deviations of these data from the best-fit great circle are very small ($<0.4^\circ$).

From the temporal and spatial constraints, we conclude that Saint Martin, Manicouagan and Rochechouart were generated by projectiles that were essentially coaxial with respect to each other (like Shoemaker–Levy 9; refs 24, 25). The spatial relations between the projectiles that generated Obolon' and Red Wing and those that generated the three larger impacts are not clear. Red Wing, at only 9 km diameter, could have been produced by a fragment of the same projectile that generated the larger Saint Martin structure.

It is probable that there were more than five impact structures generated by the fragmented bolide. Those craters generated by fragments that hit the Tethys or Panthalassa oceans rather than Pangaea, however, would have been subsequently destroyed by subduction. It is of interest to note that two other impact structures show possible spatial associations with the five already discussed. Wells Creek, Tennessee, USA (12 km diameter, 200 ± 100 Myr in age)³, lies on a great circle that intersects Manicouagan, with a declination identical to that of the two great circles shown in Fig. 2 (37.2°). Newporte, North Dakota, USA (3 km diameter, <500 Myr in age)³, lies on the same great circle as, and between, Red Wing and Saint Martin. However, the age constraints for these two structures will need refining before any association with the five principal

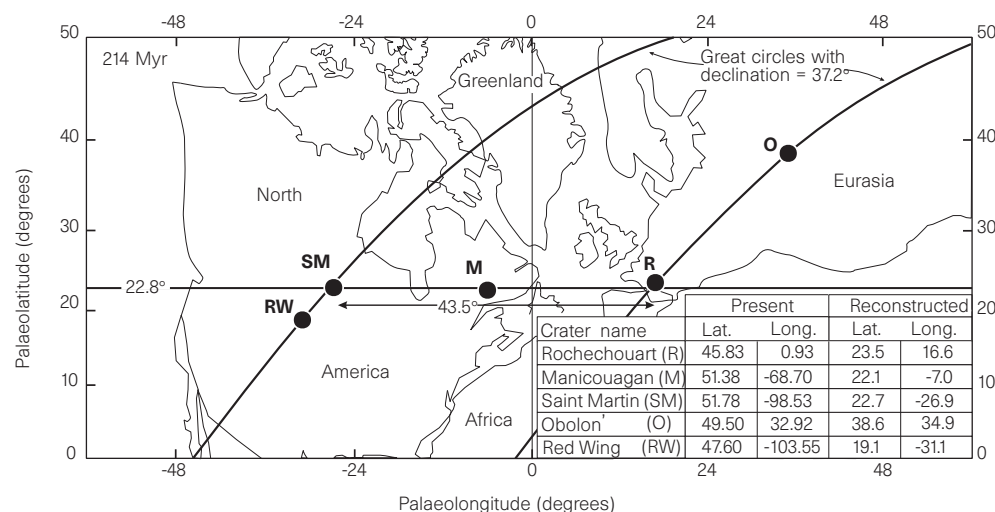


Figure 2 Reconstruction of the North American (Laurentian) and Eurasian plate positions in the Northern Hemisphere of Earth 214 Myr ago (Mercator projection), showing the locations of the five impact structures.

structures can be considered further.

If the five impact structures are contemporaneous then they would have been generated by projectiles of common affinity. Attempts at identifying the body type of the impactor at Manicouagan and Saint Martin have been unsuccessful; melt sheet analyses yield low to undetectable iridium values²⁶. Initial work on Rochechouart melt rocks implied contamination by a IIA iron meteorite²⁷, but subsequent analysis of footwall metal veins revealed an elevated Cr content and Ni/Co ratios more compatible with a chondritic source²⁸. Small amounts of kamacite and taenite have been reported in the heavy mineral fractions of allogenic breccia from Obolon¹⁵, but their presence does not decisively distinguish an iron from a chondritic source. No appraisal has yet been made of Red Wing. Melt sheets may be poor indicators of projectile compositions owing to the effects of crustal dilution, and depending on the velocity of impact and subsequent mode of projectile dissemination²⁹. It should also be cautioned that the non-volatile components of comets have chemical compositions close to those of primitive carbonaceous chondrites³⁰. Furthermore, because comets comprise roughly equal portions of silicates, hydrocarbons and volatile ices, their expected bulk siderophile contents are expected to be less than one-third that of an undifferentiated carbonaceous asteroid. This may make detection of a cometary projectile difficult. More work is needed to identify the projectile for the five impact structures. Until then, it is not possible to determine whether the source was a comet or an asteroid on the basis of impact structure geochemistry alone.

The Norian stage of the Triassic is marked by one or more major mass extinction events. One of these occurred at the Triassic/Jurassic boundary at ~205 Myr (refs 12, 22, 31) and is associated with impact-generated shocked-quartz-bearing shale beds in Italy³². This is significantly later than the 214-Myr multiple impact event described here. However, Benton³³ has alluded to an earlier Carnian-Norian mass extinction event at ~220 Myr, which affected tetrapods and plants on land and certain marine groups³⁴. Unlike the end-Triassic, the age of the Carnian/Norian boundary is less well constrained and it is possible that a 214-Myr multiple impact could be associated with an extinction event at that time. As the late Triassic period is now known to contain evidence for a multiple impact event, a renewed effort should be made to study Carnian/Norian boundary and Norian sections with the aim of locating and characterizing the impact ejecta layer, as well as refining biostratigraphic-radiometric correlations for that time period.

The calculated probability of a multiple impact occurring on Earth, comparable to that of Shoemaker-Levy 9 colliding with Jupiter, appears to be low³⁵. This is because, relative to Earth, Jupiter's huge mass presents a formidable gravitational attraction and disruptive force for asteroids and comets. Pairs of craters formed by near-simultaneous impact of binary asteroids are expected for ~10% of the impact structures on Earth and Venus³⁶. However, a terrestrial multiple impact event resulting from the collision of several projectiles would appear to require a special situation, possibly involving the close fly-by, orbital capture and decay, and consequent clustered strike of a fragmented or weakly coherent comet or asteroid³⁵. The mechanisms by which such a particular circumstance may arise now need to be further explored. Critical to identifying any additional crater chains on Earth will be the continued precision dating of proven impact structures, followed by their placing on accurate plate reconstructions for the time of impact. □

Received 30 January; accepted 19 February 1998.

- Orton, G. *et al.* Collision of comet Shoemaker-Levy 9 with Jupiter observed by the NASA infrared telescope facility. *Science* **267**, 1277–1282 (1995).
- Melosh, H. J. & Schenk, P. Split comets and the origin of crater chains on Ganymede and Callisto. *Nature* **365**, 731–733 (1993).
- Grieve, R. A. E., Rupert, J., Smith, J. & Theriault, A. The record of terrestrial impact cratering. *GSA Today* **5**, 189–196 (1995).

- Alvarez, L. W., Alvarez, W., Asaro, F. & Michel, H. V. Extraterrestrial cause for the Cretaceous-Tertiary extinction. *Science* **208**, 1095–1108 (1980).
- Hildebrand, A. R. *et al.* Chicxulub crater: a possible Cretaceous/Tertiary boundary impact crater on the Yucatán Peninsula. *Geology* **19**, 867–871 (1991).
- Sharpton, V. L. & Ward, P. E. (eds) Global catastrophes in earth history: an interdisciplinary conference on impacts, volcanism and mass mortality. *Spec. Pap. Geol. Soc. Am.* **247** (1990).
- Lambert, P. The Rochechouart crater: shock zoning study. *Earth Planet. Sci. Lett.* **35**, 258–268 (1977).
- Lambert, P. *La structure d'impact de météorite géante de Rochechouart*. Thesis, Univ. Paris-Sud (1974).
- Reimold, W. U. & Oskierski, W. in *Research in Terrestrial Impact Structures* (ed. Pohl, J.) 94–114 (Vieweg, Braunschweig/Weissbaden, Germany, 1987).
- Kelley, S. P. & Spray, J. G. A late Triassic age for the Rochechouart impact structure, France. *Meteorit. Planet. Sci.* **32**, 629–636 (1997).
- Dworak, U. Stoßwellenmetamorphose des Anorthosits vom Manicouagan Krater, Québec, Canada. *Contrib. Mineral. Petrol.* **24**, 306–347 (1969).
- Hodych, J. P. & Dunning, G. R. Did the Manicouagan impact trigger end-of-Triassic mass extinction? *Geology* **20**, 51–54 (1992).
- McCabe, H. R. & Bannatyne, B. B. Lake St. Martin cryptoexplosion crater and geology of surrounding area. *Geol. Surv. Manitoba Geol. Pap.* **3/70** (1970).
- Reimold, W. U., Barr, J. M., Grieve, R. A. F. & Durrheim, R. J. Geochemistry of the melt and country rocks of the Lake St. Martin impact structure, Manitoba, Canada. *Geochim. Cosmochim. Acta* **54**, 2093–2111 (1990).
- Masaitis, V. L., Danilina, A. N., Karpov, G. M. & Raykhlin, A. I. Karla, Obolon' and Rotmistrovka astroblemes in the European part of the U.S.S.R. *Dokl. Akad. Nauk SSSR* **230**, 174–177 (1976).
- Gurov, Y. P., Val'ter, A. A. & Rakitskaya, R. B. Coesite in rocks of meteorite explosion craters on the Ukrainian shield. *Int. Geol. Rev.* **22**, 329–332 (1978).
- Gurov, Y. P., Gurova, E. P. & Rakitskaya, R. B. Impact diamonds in the craters of the Ukrainian shield. *Meteoritics* **30**, 515–516 (1995).
- Masaitis, V. L. *et al.* *The Geology of Astroblemes* (Nedra Press, Leningrad, 1980).
- Grieve, R. A. F. & Masaitis, V. L. The economic potential of terrestrial impact craters. *Econ. Geol.* **36**, 105–151 (1994).
- Gerhard, L. C., Anderson, S. B., Lefever, J. A. & Carlson, C. G. Geological development, origin, and energy mineral resources of Williston Basin, North Dakota. *Am. Assoc. Petrol. Geol. Bull.* **66**, 989–1020 (1982).
- Melosh, H. J. *Impact Cratering: A Geologic Process* (Oxford Univ. Press, 1989).
- Gradstein, F. M. *et al.* A Mesozoic time scale. *J. Geophys. Res.* **99**, 24051–24074 (1994).
- Ziegler, A. M. *et al.* in *The Tectonic evolution of Asia* (eds Yin, A. & Harrison, T. M.) 371–400 (Cambridge Univ. Press, 1996).
- Hammel, H. B. *et al.* HST imaging of atmospheric phenomena created by the impact of comet Shoemaker-Levy 9. *Science* **267**, 1288–1296 (1995).
- Weaver, H. A. *et al.* The Hubble Space Telescope (HST) observing campaign on comet Shoemaker-Levy 9. *Science* **267**, 1282–1288 (1995).
- Palme, H. Identification of projectiles of large terrestrial impact craters and some implications for the interpretation of Ir-rich Cretaceous/Tertiary boundary layers. *Geol. Soc. Am. Spec. Pap.* **190**, 223–233 (1982).
- Jannsens, M. J., Hertogen, J., Takahashi, H., Anders, E. & Lambert, P. Rochechouart impact crater: identification of projectile. *J. Geophys. Res.* **82**, 750–758 (1977).
- Horn, N. P. & Goresy, A. E. The Rochechouart crater in France: stony and not iron meteorite? *Lunar Planet. Sci.* **11**, 468–470 (1980).
- McClaren, D. J. & Goodfellow, W. D. Geological and biological consequences of giant impacts. *Annu. Rev. Earth Planet. Sci.* **18**, 123–171 (1990).
- Hut, P. *et al.* Comet showers as a cause of mass extinctions. *Nature* **329**, 118–126 (1987).
- Hallam, A. in *Global Events and Event Stratigraphy* (ed. Walliser, O. H.) 265–283 (Springer, New York, 1996).
- Bice, D. M., Newton, C. R., McCauley, S., Reiners, P. W. & McRoberts, C. A. Shocked quartz at the Triassic/Jurassic boundary in Italy. *Science* **259**, 443–446 (1992).
- Benton, M. J. More than one event in the late Triassic mass extinction. *Nature* **321**, 857–861 (1986).
- Hallam, A. & Wignall, P. B. *Mass Extinctions and their Aftermath* (Oxford Univ. Press, 1997).
- Love, S. G., Botke, W. F. & Richardson, D. C. Alternative formation mechanisms for terrestrial crater chains. *Lunar Planet. Sci.* **XXVIII**, 837–838 (1997).
- Botke, W. F. & Melosh, H. J. Formation of asteroid satellites and doublet craters by planetary tidal forces. *Nature* **381**, 51–53 (1996).

Acknowledgements. We thank M. Benton and J. Melosh for discussions and reviews. This work was supported by NSERC (Canada) and the Open University (UK) grants to J.G.S. and S.P.K., respectively. D.B.R. acknowledges industrial support for the Palaeogeographic Atlas Project.

Correspondence and requests for materials should be addressed to J.G.S. (e-mail: jgs@unb.ca).

Fission-track ages of stone tools and fossils on the east Indonesian island of Flores

M. J. Morwood*, P. B. O'Sullivan†, F. Aziz‡ & A. Raza†

* Department of Archaeology and Palaeoanthropology, University of New England, New South Wales 2351, Australia

† School of Earth Sciences, La Trobe University, Victoria 3083, Australia

‡ Geological Research and Development Centre, Bandung 4011, Indonesia

The islands of Wallacea, located between the Southeast Asian (Sunda) and Australian (Sahul) continental areas, offer unique potential for the study of evolution and cultural change. Located east of Java and Bali, which were periodically connected to the Asian mainland, the Wallacean islands could only be reached by

sea crossings. Consequently, before human intervention all these islands had impoverished faunas comprising only species that were capable of crossing water by swimming, rafting on flotsam, or by flying in sufficient numbers to establish biologically viable populations¹. Here we report zircon fission-track dates from two fossil sites on the Wallacean island of Flores. Tangi Talo, which has an endemic fauna, dates to 0.90 ± 0.07 Myr BP, whereas Mata Menge, where stone tools are found with elements of continental Southeast Asian fauna, dates to between 0.88 ± 0.07 and 0.80 ± 0.07 Myr BP. Even at times when the sea level was lowest, water crossings were necessary to reach Flores from Southeast Asia. We conclude that *Homo erectus* in this region was capable of repeated water crossings using watercraft.

Tangi Talo and Mata Menge are two of many palaeontological sites located in the Ola Bula Formation of the upper Ae Sissa River basin in central Flores (Fig. 1). The Ola Bula Formation is up to 80 m thick and comprises white tuffs and interbedded tuffaceous sediments underlying fluvial deposits. On the basis of stratigraphic context and faunal content, Mata Menge appears to be younger than the sites of Boa Leza and Ola Bula. Tangi Talo lies 31 m stratigraphically below Ola Bula, and is therefore older than Mata Menge².

At Tangi Talo, a horizontally bedded white tuffaceous layer ~30 cm thick contains the remains of pygmy *Stegodon*, giant tortoises (*Geochelone* sp.) and Komodo dragons (*Varanus komodoensis*). Although angular volcanic rock fragments are locally present, none of these remains have characteristics that suggest

they are artefacts. Palaeomagnetic determinations showed that deposits at Tangi Talo are dominated by reversed polarities but contain two levels with normal polarities. These two levels were interpreted to represent the Jaramillo normal polarity subzone with a date of ~900,000 years BP (ref. 2); however, they could also have resulted from the inefficiency of the field demagnetization procedures used on the samples for removing normal overprints held in secondary minerals common in tropical weathering (ref. 2 and C. Swisher, personal communication).

The Mata Menge site comprises an ~120-cm-thick section of horizontally bedded, white tuff with interbedded, tuffaceous sandstones (Fig. 2, Unit B). The section is interpreted here in terms of the primary, or near primary, deposition of tuff with intruded tuffaceous flow deposits. In places, soft-sediment deformation is evident between layers, suggesting that the units were subaqueously deposited in a low-energy lacustrine environment. The tuff and its contents therefore mainly accumulated from above, whereas the sandstones and their contents entered laterally. These deposits contain the remains of large *Stegodon* (*S. trigonocephalus florensis*), crocodile, giant rat (*Hooijeromys nusatenggara*), freshwater molluscs, plants and rounded volcanic pebbles.

In addition, the Mata Menge deposits contain pieces of volcanic rock and chert identified as artefacts on the basis of well-defined flake scars, ring cracks, bulbs of percussion and systematic edge damage suggestive of retouch (Fig. 3) (refs 3,4). The chert flakes are also out of geological context in the volcanic deposits of the Ola Bula Formation and lack any evidence of being transported by water. Of 45 stone pieces recovered in a 1994 excavation (ref. 5), 14 were identified as artefacts according to technological criteria³. Of these 14 artefacts, four were subsequently examined under high magnification and found to have edge damage, striations, polishing and residues indicating use in the processing of plant material³. Further *in situ* stone artefacts were located in the tuffaceous flow deposits (but not in the white tuff) at Mata Menge during fieldwork early in 1997. The artefacts included a chert flake at the base of the fossil deposit, just 3 cm above the lowermost dated tuff sample. The flake has bifacial usewear, striations and residues.

The fossil layer at Mata Menge is overlaid by another 23 m of the Ola Bula Formation capped by lacustrine limestone⁶. Immediately below the fossil layer are extensive layers of tuff and sandstone, which contain neither fossils nor stone artefacts. Palaeomagnetic determinations undertaken previously found a reversed-normal transition in deposits 3 m below the fossil-rich layer at Mata Menge. The transition was interpreted as representing the Matuyama/Brunhes boundary, and a date of slightly less than

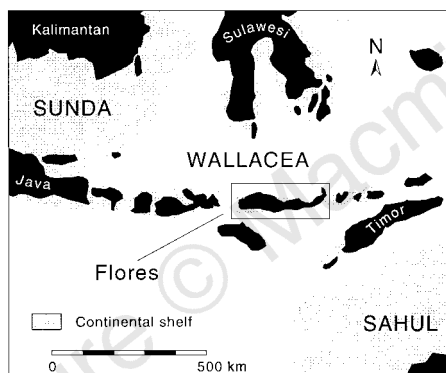


Figure 1 The general location of Flores in east Indonesia. The continental shelves would have been exposed to varying water levels at times of glacial maximum/low sea level. Black areas represent land masses that are now above water.

Table 1 Fission track analytical results: tuffaceous samples from Flores, Indonesia

Sample number (no. of grains)	Lat. (°S) Long. (°E)	Unit name	Standard track density* $\times 10^5 \text{ cm}^{-2}$	Fossil track density* $\times 10^5 \text{ cm}^{-2}$	Induced track density* $\times 10^5 \text{ cm}^{-2}$	Chi-squared probability (%)	Uranium (p.p.m.)	Fission-track age (Myr $\pm$ s.d.)
Primary zircon: Mata Menge								
MM1	8° 41' 31"	Ola Bula	4.060	1.139	3.311	33.1	318.0	0.88 ± 0.07
(44)	121° 5' 43"		(1,631)	(199)	(5,784)			
MM2	8° 41' 31"	Ola Bula	4.087	1.040	3.340	15.1	318.7	0.80 ± 0.07
(53)	121° 5' 43"							
Primary zircon: Tangi Talo								
TT1	8° 41' 53"	Ola Bula	4.114	1.195	3.441	100.0	326.2	0.90 ± 0.07
(56)	121° 8' 10"		(1,631)	(184)	(5,297)			
Detrital contamination: Mata Menge								
MM1	-	-	-4.060	9.171	1.766	0.0	169.7	9.8 ± 3.1
(6)	-	-	(1,631)	(229)	(441)			
MM2	-	-	-4.087	4.200	1.642	0.1	156.6	7.5 ± 1.7
(7)	-	-	(1,631)	(98)	(383)			
Detrital contamination: Tangi Talo								
TT1	-	-	-4.114	0.136	1.312	0.0	124.4	14.0 ± 8.3
(4)	-	-	(1,631)	(89)	(86)			

* Brackets show number of tracks counted. Standard and induced track densities were measured on mica external detectors ($g = 0.5$) and fossil track densities on internal mineral surfaces. Zircon ages were calculated using $t = 125.5 \pm 3$ for dosimeter glass CN1 (analyst, P.B.O.'S.).

0.73 Myr BP for the faunal remains was inferred², although a date of about 0.78 Myr BP for the boundary event is more likely⁷. The palaeomagnetic work did not date the fossil deposits themselves, however.

To constrain better the age of the deposits at Mata Menge and Tangi Talo, we collected samples from three tuffaceous horizons for

zircon fission-track dating. This method of radiometric dating is ideal for dating young tuffaceous deposits⁸. The stratigraphic positions of the two samples taken from Mata Menge (MM) are shown in Fig. 2. Sample MM1 was collected from the pink tuff immediately below the lowest fossils and stone artefacts identified, whereas sample MM2 was taken from the white tuff next to a large

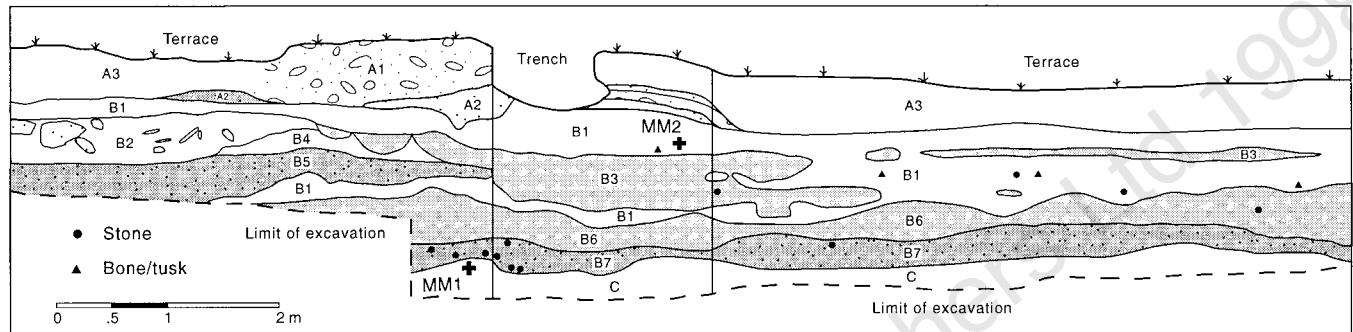


Figure 2 The west bank of the excavations at Mata Menge, showing the locations of fission-track samples MM1 and MM2 (marked by plus symbols). Three main stratigraphic units are evident. Unit A is a layer of weathered sandstone and top soil containing stone artefacts. A1 is gravel-rich; A2 is an *in situ* homogeneous, coarse sandstone; A3 is top soil. Unit B comprises layers and lenses of white tuff and tuffaceous sandstone flow structures, which predominate towards the base. The flow structures contain more fragmented fossils and

all stone artefacts identified in the section. B1 is a white tuff; B2 is a light-grey tuffaceous sandstone; B3 is a dark-grey tuffaceous sandstone; B4 is a light-grey tuffaceous sandstone; B5 is a brown-grey sandstone; B6 is a dark-grey sandstone (similar to B3 but with inclusions); B7 is a coarse-grained, yellow-grey sandstone. Unit C is a pink tuff, which does not contain fossils or stone artefacts. The section around the trench is cut back 50 cm from the main face.

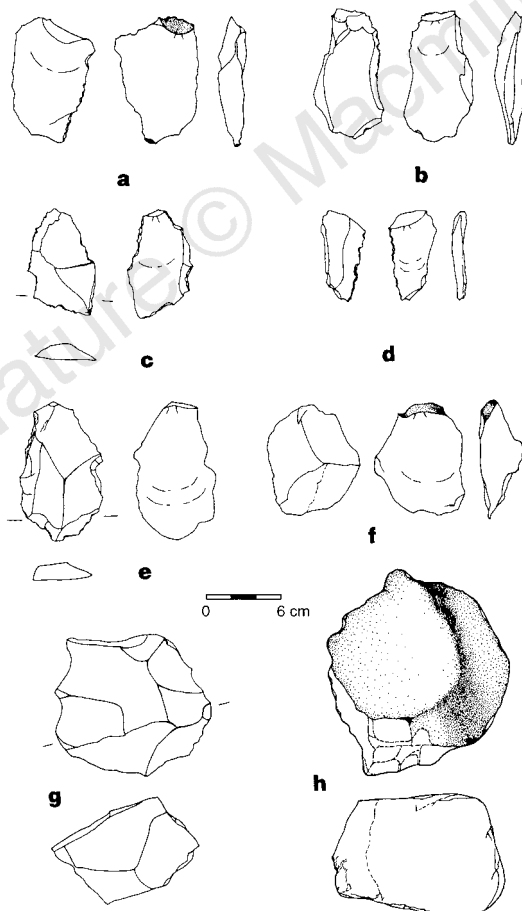


Figure 3 Stone artefacts excavated from Mata Menge in 1994. **a**, Chert flake; **b**, silcrete flake; **c-f**, basalt flakes; **g**, basalt multi-platformed core; **h**, basalt cobble 'chopper'.

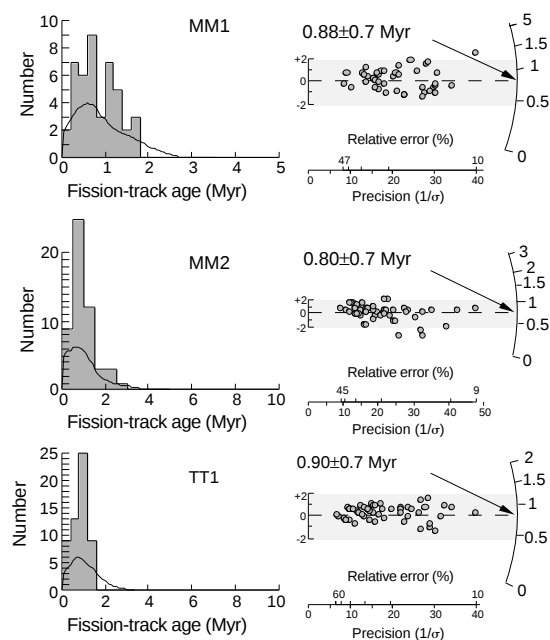


Figure 4 Single-grain age results in the form of simple histograms (left) and radial plots (right) from the three Mata Menge and Tangi Talo tuffaceous units MM1, MM2 and TT1. On radial plots, each age has unit standard error ($\pm 2\sigma$) on the y-axis; its actual precision is indicated on the x-axis; and its age is read by extrapolating a line from zero through the plotted point to the logarithmic age scale on the right perimeter.

Stegodon tusk. It overlies *in situ* artefacts. Another sample, TT1, was taken from the white tuffaceous layer at Tangi Talo next to an *in situ* pygmy *Stegodon* tusk and pieces of *Geochelone* carapace. Although the tuffaceous layers sampled may have been slightly reworked since deposition, their horizontal extent as well as homogeneity in colour and texture suggest that any such reworking has been minimal. Therefore, the measured ages of zircons from these units should represent the depositional age of the tuffaceous layers.

Analytical results are shown in Table 1. Distributions of the primary grain ages from each sample are presented in Fig. 4. The Tangi Talo sample (TT1) yielded a date of 0.90 ± 0.07 Myr BP, whereas the two Mata Menge samples (MM1 and MM2) yielded dates of 0.88 ± 0.07 and 0.80 ± 0.07 Myr BP, respectively. These closely match the dates previously suggested on the basis of palaeomagnetic determination², biostratigraphy⁴ and the presence of tektites⁹.

We conclude that at Mata Menge stone artefacts occur in primary association with Early Pleistocene tuffs and faunal remains. The age of the artefacts also indicates that they were produced by *Homo erectus* rather than *Homo sapiens*¹⁰.

There are several general implications of these results. For instance, today three deep-water straits occur between the Sunda continental shelf and Flores. Even at times of lowest sea level during the last glacial maximum, these straits required sea crossings of at least 19 km (ref. 11). Either there must have been a land bridge linking Flores with mainland Southeast Asia in the Early Pleistocene epoch, or *Homo erectus* in this region had the capacity to make water crossings¹². The impoverished nature of the fauna on Flores throughout the sequence, however, seems to negate a connection with Sunda at any time (compare refs 12, 13). The presence of endemic pygmy elephants, giant reptiles and giant rats in the Early Pleistocene also suggests a continued insular context^{14,15} both before and contemporaneous with the first evidence for hominids. In fact, Sondaar¹² has argued that the extinction of pygmy *Stegodon* and giant tortoises, followed by recolonization of Flores by large *Stegodon* from continental Southeast Asia after 0.90 Myr BP, resulted from increased predation, not reduced insularity. More specifically, he says that the onset of hunting by hominids caused the turnover in fauna, a position supported by our results.

Our findings, therefore, validate generally dismissed (or ignored) claims for the presence of hominids on Flores in the Early/Middle Pleistocene (compare refs 2, 4 with 16, 17). Furthermore, they indicate that, sometime between 800,000 and 900,000 years ago, *Homo erectus* in this region had acquired the capacity to make water crossings. Previously, in the region of the Wallacean islands this capacity was thought to be the prerogative of modern humans and to have only appeared in the Late Pleistocene, with the earliest widely accepted evidence for watercraft being the colonization of Australia by modern humans from Wallacea between 40,000 and 60,000 years ago^{18,19}. Outside this region, the technology to undertake even limited water crossings is not clearly evident until much later, at the end of the Pleistocene²⁰. Our results challenge this view. Therefore, this evidence, combined with the geographical radiation of *Homo erectus* in the Early Pleistocene¹⁰ and other recent discoveries²¹, suggests that the cognitive capabilities of this species may be due for reappraisal. □

Methods

Tuffaceous samples were crushed and zircons in the size range 60–250 µm were separated using a Gemini table and standard magnetic and heavy-liquid techniques at the School of Earth Sciences, La Trobe University. Grains were mounted in FEP Teflon, and, after optical-quality polishing, were etched in a eutectic KOH/NaOH melt^{7,22} at 225 °C for ~125 h. Neutron irradiations were carried out in a well-thermalized flux in the HIFAR reactor (Lucas Heights, Australia). Fission-track ages were measured using the external detector method²³ with Brazil Ruby muscovite being used to record induced tracks. The muscovite detectors were etched for 30 min in 48% hydrogen fluoride at

room temperature to reveal the induced tracks. Thermal neutron fluences were monitored by measuring the track density recorded in muscovites attached to pieces of the Corning CN1 standard glass. Fission tracks in each zircon mount were counted in transmitted light using a dry ×100 objective lens at a total magnification of ×1,600. Only grains with sharp polishing scratches were counted. Depending on availability of datable grains, a minimum of 50–60 individual grains were dated from each sample. Ages were calculated using the standard fission-track-age equation using the zeta calibration method²⁴ and errors were calculated using previously published methods²⁵. A personal zeta calibration factor of 125.5 ± 3 (for P.B.O.'S) was determined empirically using zircon-age standards with independently known ages.

The use of the external detector method was particularly important as it allowed dating of individual grains within a sample, rather than relying on a bulk sample age which can contain contaminants. The χ^2 statistic was used to detect the probability that all age grains analysed belonged to a single population of ages. All three mounts showed a probability of <5%, which was evidence of asymmetric spreads of single-grain ages. Therefore, in each case the 'central age' (ref. 26), which is essentially a weighted-mean age, was reported. Data from grains deemed on the basis of their significantly older ages to be obvious contaminants were removed before a final age calculation was made. Each sample contained between four and seven older reworked zircons, which gave ages of between 7 and 14 Myr BP, much older than the primary zircons. These older contaminants also contained significantly lower amounts of uranium (Table 1), a factor that could also have been used to delineate between the primary and contaminate zircons.

Received 27 June; accepted 20 November 1997.

1. Johnson, D. L. Problems in the land vertebrate zoogeography of certain islands and the swimming powers of elephants. *J. Biogeog.* **7**, 383–398 (1980).
2. Sondaar, P. Y. *et al.* Middle Pleistocene faunal turnover and colonisation of Flores (Indonesia) by *Homo erectus*. *C. R. Acad. Sci.* **319**, 1255–1262 (1994).
3. Morwood, M. J., Aziz, F., van den Bergh, G. D., Sondaar, P. Y. & de Vos, J. Stone artefacts from the 1994 excavation at Mata Menge, west central Flores, Indonesia. *Austr. Archaeol.* **44**, 26–34 (1997).
4. Maringer, J. & Verhoeven, T. Die Steinartefakte aus der *Stegodon*—Fossilischiicht von Mengeruda auf Flores, Indonesien. *Anthropos* **65**, 229–247 (1970).
5. van den Bergh, G. D., Mubrot, B., Aziz, F., Sondaar, P. Y. & de Vos, J. In *Indo-Pacific prehistory: the Chiang Mai Papers* (Proc. 15th IPPA Congr. Chiang Mai, Thailand 1994 Vol. 1) (ed. Bellwood, P.) 27–36 (IPPA, Canberra, 1996).
6. Lumbanbatu, U. M. & Aziz, F. Aspek fosil vertebrata dari Forasi Olabula, Flores. *J. Geol. Sumberdaya Miner.* **33**, 2–6 (1994).
7. Berggren, W. A. *et al.* in *Geochronology, Time Scales and Global Stratigraphic Correlation* (eds Berggren, W. A. *et al.*) 129–212 (SEPM, Special publ. no. 54, Tulsa, 1995).
8. Gleadow, A. J. W. Fission track ages of the KBS tuff and associated hominid remains in northern Kenya. *Nature* **284**, 225–230 (1980).
9. von Koenigswald, G. H. R. A tektite from the island of Flores (Indonesia). *Proc. Koninklijke Nederlandse Akad. Wetenschappen B* **61**, 44–46 (1957).
10. Rightmire, G. P. in *100 Years of Pithecanthropus: the Homo erectus Problem* (ed. Franzen, J. L.) 319–326 (Courier Forschungsinstitut Senckenberg, Frankfurt, 1994).
11. Birdsell, J. H. in *Sunda and Sahul* (eds Allen, J. *et al.*) 113–168 (Academic, London, 1977).
12. Sondaar, P. Y. Pleistocene man and extinctions of island endemics. *Mem. Soc. Geol. Fr.* **150**, 159–165 (1987).
13. Groves, C. P. in *Perspectives in Human Biology*, Vol. 2 (eds Rousham, E. & Freedman, L.) 83–87 (Univ. Western Australia, Perth, 1996).
14. Davis, S. Tiny elephants and giant mice. *New Sci.* **105**, 25–27 (1985).
15. Azzaroli, A. About pygmy mammoths of the Northern Channel Islands and other island faunas. *Quat. Res.* **16**, 423–425 (1981).
16. Allen, H. *Stegodonts and the dating of stone tool assemblages in island S.E. Asia*. *Asian Persp.* **30**, 243–266 (1991).
17. Bellwood, P. *Prehistory of the Indo-Malaysian Archipelago* (Academic, Sydney, 1985).
18. Davidson, I. & Noble, W. Why the first colonisation of the Australian region is the earliest evidence of modern human behaviour. *Archaeol. Oceania* **27**, 135–142 (1992).
19. Bowdler, S. in *Sahul in Review* (eds Smith, M. A. *et al.*) 60–70 (Australian National University, Canberra, 1993).
20. Cherry, J. F. The first colonisation of the Mediterranean Islands. *J. Med. Archaeol.* **3**, 145–221 (1990).
21. Thieme, H. Lower Palaeolithic hunting spears from Germany. *Nature* **385**, 807–810 (1997).
22. Gleadow, A. J. W., Hurlford, A. J. & Quaipe, D. R. Fission track dating of zircon: improved etching techniques. *Earth Planet. Sci. Lett.* **33**, 273–276 (1976).
23. Gleadow, A. J. W. Fission track dating methods: what are the real alternatives? *Nucl. Tracks* **5**, 3–14 (1981).
24. Hurlford, A. J. & Green, P. F. A users' guide to fission-track dating calibration. *Earth Planet. Sci. Lett.* **59**, 343–354 (1982).
25. Green, P. F. A new look at statistics in fission-track dating. *Nucl. Tracks* **5**, 77–86 (1981).
26. Galbraith, R. F. & Laslett, G. M. Statistical models for mixed fission track ages. *Nucl. Tracks* **21**, 459–470 (1993).

Acknowledgements. Fieldwork conducted between 1991 and 1994 by P. Sondaar, J. de Vos, G. van den Bergh and F.A. provided the basis for this investigation. Our 1997 fieldwork was undertaken under the auspices of the Indonesian Geological Research and Development Centre and was funded by a grant from the University of New England Vice Chancellor's Fund (we thank B. Thom for this support), and by the Australian Institute of Nuclear Science and Engineering. We thank V. Paine for help with stratigraphic recording; R. Spikings for completing the mineral separations; D. Hobbs and K. Morwood for artwork; and I. Davidson, J. de Vos, P. Sondaar, D. Cogill, P. Brown, R. Gargett, C. Swisher, P. Bellwood and R. Roberts for discussion.

Correspondence and requests for materials should be addressed to M.J.M. (e-mail: mmorwood@metz.une.edu.au).

Regulation of ovulation by human pheromones

Kathleen Stern & Martha K. McClintock

Department of Psychology, The University of Chicago, 5730 Woodlawn Ave, Chicago, Illinois 60637, USA

Pheromones are airborne chemical signals that are released by an individual into the environment and which affect the physiology or behaviour of other members of the same species¹. The idea that humans produce pheromones has excited the imagination of scientists and the public, leading to widespread claims for their existence, which, however, has remained unproven. Here we investigate whether humans produce compounds that regulate a specific neuroendocrine mechanism in other people without being consciously detected as odours (thereby fulfilling the classic definition of a pheromone). We found that odourless compounds from the armpits of women in the late follicular phase of their menstrual cycles accelerated the preovulatory surge of luteinizing hormone of recipient women and shortened their menstrual cycles. Axillary (underarm) compounds from the same donors which were collected later in the menstrual cycle (at ovulation) had the opposite effect: they delayed the luteinizing-hormone surge of the recipients and lengthened their menstrual cycles. By showing in a fully controlled experiment that the timing of ovulation can be manipulated, this study provides definitive evidence of human pheromones.

The existence of human pheromones was first suggested by the demonstration that women living together can develop synchronized menstrual cycles under specific conditions²⁻⁵. In rats, a similar process of ovarian synchrony occurs and is mediated by the exchange of two different pheromones⁶⁻⁷. One, produced before ovulation, shortens the ovarian cycle; the second, produced at ovulation, lengthens the cycle. These two opposing pheromones were predicted by a coupled-oscillator model of ovarian synchrony and shown by computer simulation to be sufficient for producing not only synchrony, but also the other observed effects of ovarian asynchrony and cycle stabilization^{7,8}. By applying this model to humans, we demonstrate the existence of human pheromones and

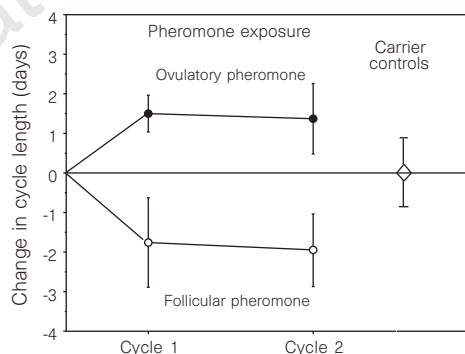


Figure 1 Effect of axillary compounds, donated by women during the follicular or ovulatory phases of their menstrual cycle, on the menstrual cycle length of recipients. This was measured as a change in length from the recipient's baseline cycle with a repeated measures analysis of variance: within-subject factors were follicular versus ovulatory compounds ($F(1,18) = 5.81, P \leq 0.03$) and cycle 1 versus cycle 2 of exposure (not significant, NS); the between-subjects factor was: order of presentation (NS); all interactions between factors were not significant. Cycles were shorter than baseline during exposure to follicular compounds ($t = 1.78, P \leq 0.05$, 37 cycles) but longer during exposure to ovulatory compounds ($t = 2.7, P \leq 0.01$, 38 cycles). Cycles during exposure to the carrier were not different from baseline ($t = 0.05, P \leq 0.96$, 27 cycles).

identify a potential pheromonal mechanism for menstrual synchrony, as well as for other forms of social regulation of ovulation.

We found that the recipients had shorter cycles when receiving axillary compounds produced by donors in the follicular phase of the menstrual cycle (-1.7 ± 0.9 days) and longer cycles when receiving ovulatory compounds ($+1.4 \pm 0.5$ days), which represent significantly different opposite effects (Fig. 1). The response was manifest within the first cycle, rather than requiring three cycles of exposure as suggested previously^{2,7}, and the sequence of compound presentation had no effect. The two types of axillary compounds had effects that were significantly different from each other and from the baseline cycle. The carrier had no effect on cycle lengths of the control recipients. In five of the cycles, women had mid-cycle nasal congestion, which could have prevented their exposure to pheromones; including these cycles in the analysis made the results slightly less robust (follicular compounds: -1.4 ± 0.9 days; ovulatory compounds: $+1.4 \pm 0.5$ days; ANOVA: follicular versus ovulatory compounds $F(1,18) = 4.32, P \leq 0.05$; cycle 1 versus

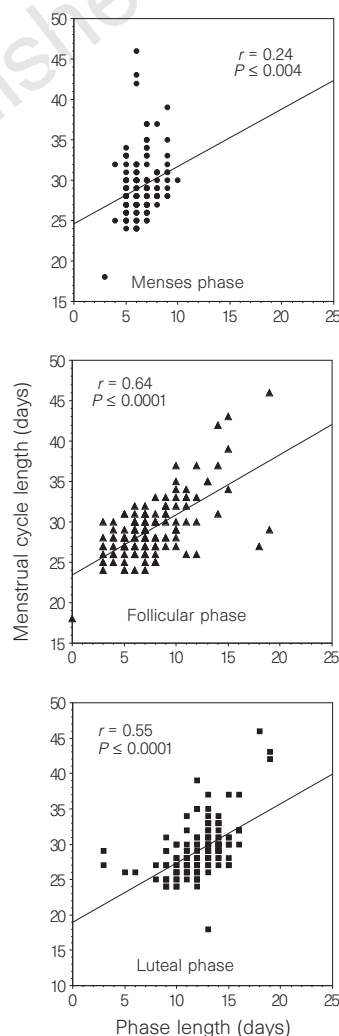


Figure 2 Each of the three phases of the menstrual cycle are variable in length (x-axis) and correlate with overall menstrual cycle length (Pearson's r), establishing each as a potential mediator of the effects of axillary compounds. Menses phase (●, day 1 to the end of menses); follicular phase (▲, day after menses to the day before the preovulatory LH surge); luteal phase (■, three days after the LH surge to day before menses, verified to be functional by ovulatory levels of progesterone glucuronide (PG) and rise in basal body temperature. The ovulatory phase is a fixed 3 day interval (day of LH surge onset plus 2 subsequent days). Note that the luteal phase of these normal subjects is significantly more variable than the 12–16-day range described in standard medical texts.

cycle 2 of exposure (not significant, NS); order of presentation (NS); all interactions between factors were not significant).

The finding that axillary compounds changed cycle length indicates that the compounds contain pheromones. The existence of two opposing effects, especially one that accelerates ovulation, makes it unlikely that a simple disruption of ovulation by a chemical produced the observed changes⁹. It also suggests two functionally different ovarian-dependent pheromones in humans, as in rats, with opposing effects. The existence of a phase-advance pheromone and a phase-delay pheromone supports the coupled-oscillator model of menstrual synchrony^{7,8}. Women reported that they detected only alcohol (the control odorant, and carrier of the compounds), indicating that these changes were due to pheromones that were not consciously detected.

These results are consistent with another central prediction of the coupled-oscillator model: that there are individual differences in sensitivity to pheromones and therefore in the strength of the response⁸. Although a significant proportion of women in this experiment responded to the pheromones with changes in cycle length in the expected directions (68% of women responded to follicular pheromones, 68% to ovulatory pheromones), some women did not. In addition, the range of response magnitude was considerably more than the variation in cycle length typical for this age group¹⁰: cycles were shortened from 1 to 14 days and lengthened from 1 to 12 days.

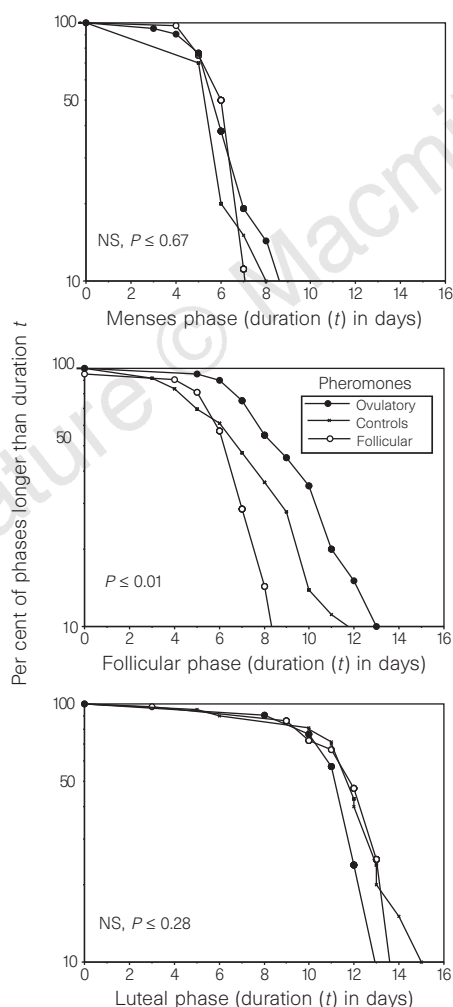


Figure 3 Effect of follicular and ovulatory pheromones on the length of each of the three phases that could mediate the observed change in menstrual-cycle length. Log-survivor analysis of the percentage of phases that are longer than a given length (time t in days; Mantel-Cox test). The same conclusions were reached with repeated measures analysis of variance on each of these three phases.

There are three phases of the menstrual cycle that vary and might mediate the effect of pheromones on cycle length; each is controlled by different neuroendocrine mechanisms (menses, follicular and luteal phases). To determine the specific mechanism of pheromone action, we measured the luteinizing hormone (LH) and progesterone glucuronide content from urine samples to pinpoint the time of the preovulatory LH surge and verify the occurrence of ovulation. Previous hypotheses have focused on the menses or luteal phases^{2-7,11,12}, although most medical texts report that the normal luteal phase is relatively fixed in length and it is the follicular phase that varies. In our sample, each of these three phases, including the luteal phase, varied significantly in length (indicated by the range of x -axis values in Fig. 2) and correlated sufficiently with cycle length for any of the three hypotheses to be correct.

Nonetheless, we traced all the changes caused by the pheromones presented in our study to the follicular phase (Fig. 3). For the menses and luteal phases, the distribution during the pheromone and control conditions were the same (indicated by overlapping log-survivor curves). Only the follicular phase was regulated, shortened by follicular compounds and lengthened by ovulatory compounds, suggesting that these ovarian-dependent pheromones have opposite effects on the recipient's ovulation by differentially altering the rate of follicular maturation or hormonal threshold for triggering the LH surge.

This experiment confirms the coupled oscillator model of menstrual synchrony and refocuses attention on the ovarian-dependent pheromones that regulate ovulation, producing either synchrony, asynchrony or cycle stabilization within a social group, namely two distinct pheromones, produced at different times of the cycle, which phase-advance or phase-delay the preovulatory LH surge. From this initial test of human ovarian-dependent pheromones, we do not know whether the phenomenon is fragile—that is, limited to modulation of ovulation timing in healthy young women—or robust, and so capable of modulating ovulation in a diverse population for either contraception or treatment of infertility. Moreover, we need to determine whether humans naturally receive compounds that have similar effects in the context of everyday life.

There may be other consequences of ovarian-dependent pheromones in women, in addition to the alteration of the timing of ovulation. Our work in rats and with computer simulations demonstrates that these same ovarian-dependent pheromones have qualitatively different effects depending on the initial conditions under which pheromonal and social interactions begin, as well as on the point in the reproductive lifespan when they occur^{7,8,13-15}. Further work in this area may well reveal that, as in rats, social interactions mediated by ovarian-dependent pheromones affect age of puberty, interbirth intervals, age at menopause, and level of chronic oestrogen exposure throughout a women's lifetime.

These data demonstrate that humans have the potential to communicate pheromonally. In other species there are many other types of pheromones, not dependent on ovarian function, which enable individuals to regulate diverse aspects of their internal neuroendocrine states on the basis of information about another's internal state or environment. For example, pheromones influence mating preference in hamsters¹⁶, dominance relationships among male elephants in musth¹⁷, timing of weaning in rats¹⁸ and how rat pups learn to distinguish edible foods from poisons¹⁹, how hamsters recognize individual members of their social group²⁰, and the level of stress experienced by a mouse in a new environment on the basis of the emotional state of the previous occupant²¹. Well controlled studies of humans are now needed to determine whether there are other types of pheromones, with effects that are as far-reaching in humans as they are in other species. □

Methods

Subjects and procedures. The experiment involved 29 women aged 20–35, who were students or staff at a university, used barrier contraception and had

histories of regular and spontaneous ovulation. They were the first women who met our subject criteria among those responding to our request for volunteers and none dropped out once the experiment had begun. We collected compounds from the axillae of 9 donor women in hormonally distinct phases of the menstrual cycle and applied them daily just under the noses of 20 recipients. All participants were unaware of the experiment's hypothesis and the source of the compounds. The study was presented as focused primarily on the development of non-invasive methods for detecting ovulation, and secondarily on sensitivity to the odour of small amounts of 'natural essences' (consent was obtained for a list of 30 compounds).

Axillary compounds. As in other species, human pheromones might be produced by apocrine glands (active only during reproductive maturity), eccrine glands (which produce sweat that contains compounds found also in saliva and urine), exfoliated epithelial cells or bacterial action^{22–24}. We collected compounds from axillae because they contain all four of these potential sources and because the two previous, albeit highly criticized, attempts to study this issue used axillary compounds^{3,4,25–28}. The 9 donors bathed without perfumed products every day and then wore 4 × 4 cotton pads in their axillae for at least eight hours. Each pad was cut into four sections for distribution to different recipients, treated with 4 drops of 70% isopropyl alcohol²⁵ and then frozen immediately at –80 °C in a glass vial.

Menstrual cycle assessment. Donors provided urine samples every evening, which we assayed for LH to detect the onset of the LH surge that triggers ovulation²⁹. This singular hormonal event unambiguously demarcates the follicular from the ovulatory phases of the cycle. The LH surge was used together with data on vaginal secretions, menses, basal body temperature, and a rise in progesterone glucuronide in the postovulatory luteal phase, to classify each pad as containing compounds produced during the follicular phase (2 to 4 days before the onset of the LH surge) or the ovulatory phase (the day of the LH surge onset and the 2 subsequent days). To ensure a similar stimulus for all recipients regardless of individual differences among donors, all 9 donors contributed equally to the follicular and ovulatory compounds received by each subject.

As it is not yet known when during the menstrual cycle women are physiologically most sensitive to putative pheromones, applying compounds every day ensured covering a potentially sensitive period. However, our computer simulation experiments indicated that in rats this pheromonal-sensitive period occurs mid-cycle, around the time of ovulation⁸ (a period when women are particularly sensitive to some olfactory stimuli³⁰). Any condition preventing exposure to the compounds, such as nasal congestion anytime during the mid-cycle period from three days before to two days after the preovulatory LH, could weaken the effect. We analysed the data taking this into account.

Experimental design. All recipients were studied for one baseline cycle without exposure to axillary compounds. Then, in a crossover experimental design during the next four consecutive cycles, axillary compounds were applied daily by wiping a thawed pad above the recipients upper lip. Half of the recipients ($n = 10$) received follicular compounds daily for two menstrual cycles and were then switched to exposure to ovulatory compounds for the next two cycles. The other 10 recipients received the same compounds in the reverse order. After applying the compounds, recipients were free to go about their normal activities but were asked not to wash their faces for the next six hours. All but two subjects, who missed only the last cycle of their second treatment, completed all five cycles of the experiment.

A between-subjects control group was provided by women (the donors) who collected all ovarian-cycle measures, but received only the carrier above their upper lip each day: 70% isopropyl alcohol. In addition, because the two-day change in menstrual cycle length (expected from the initial study²) is substantially less than individual variation in cycle length typical for this age group¹⁰, we created within-subjects controls by measuring the effect on the menstrual cycle in terms of a change in length from each individual subject's cycle preceding each condition. (For experimental subjects this was the cycle that preceded exposure to each type of compound; for control subjects this was the cycle that preceded exposure to the carrier, 70% alcohol).

Received 3 October; accepted 8 January 1998

1. Beauchamp, G. K., Doty, R. L., Moulton, D. G. & Mugford, R. A. In *Mammalian Olfaction, Reproductive Processes and Behavior* (ed. Doty, R. L.) 144–157 (Academic, New York, 1976).
2. McClintock, M. K. Menstrual synchrony and suppression. *Nature* **291**, 244–245 (1971).
3. Graham, C. A. Menstrual synchrony: an update and review. *Human Nature* **2**, 293–311 (1992).

4. Weller, L. & Weller, A. Human menstrual synchrony: A critical assessment. *Neurosci. Behav. Rev.* **17**, 427–439 (1993).
5. Weller, A. & Weller, L. Menstrual synchrony under optimal conditions: Bedouin families. *J. Comp. Psychol.* **111**, 143–151 (1997).
6. McClintock, M. K. Estrous synchrony: Modulation of ovarian cycle length by female pheromones. *Physiol. Behav.* **32**, 701–705 (1984).
7. McClintock, M. K. *Pheromones and Reproduction in Mammals* (ed. Vandenbergh, J. G.) 113–149 (Academic, New York, 1983).
8. Schank, J. & McClintock, M. K. A coupled-oscillator model of ovarian-cycle synchrony among female rats. *J. Theor. Biol.* **157**, 317–362 (1992).
9. Christian, M. S., Galbraith, W. M., Voytek, P. & Mehlman, M. A. *Advances in Modern Environmental Toxicology III: Assessment of Reproductive and Teratogenic Hazards* (Princeton Scientific, NJ, 1983).
10. Treloar, N. E., Boynton, R. E., Gorghild, G. B. & Brown, B. W. Variation of the human menstrual cycle through reproductive life. *Int. J. Fertil.* **12**, 77–126 (1967).
11. Cutler, W. B., Garcia, C. R., Kreiger, A. M. Luteal phase defects: A possible relationship between short hyperthermic phase and sporadic sexual behavior in women. *Horm. Behav.* **13**, 214–218 (1979).
12. Ryan, K. D. & Schwartz, N. B. Grouped female mice: demonstration of pseudopregnancy. *Biol. Reprod.* **17**, 578–583 (1977).
13. McClintock, M. K. Social control of the ovarian cycle and the function of estrous synchrony. *Am. Zool.* **21**, 243–256 (1981).
14. Mennella, J., Blumberg, M., Moltz, H. & McClintock, M. K. Inter-litter competition and communal nursing among Norway rats: Advantages of birth synchrony. *Behav. Ecol. Sociobiol.* **27**, 183–190 (1990).
15. Lefevre, J. L. & McClintock, M. K. Isolation accelerates reproductive senescence and alters its predictors in female rats. *Horm. Behav.* **25**, 258–272 (1991).
16. Vandenbergh, J. G. (ed.). *Pheromones and Reproduction in Mammals* (Academic, New York, 1983).
17. Rasmussen, L. E. L., Perris, T. E. & Gunawardena, R. Isolation of potential musth-alerting signals from the temporal gland secretions of male Asian elephants (*Elephas maximus*): A new method. *Chemical Senses* **19**, 540 (1994).
18. Leon, M. & Moltz, H. The development of the pheromonal bond in the albino rat. *Physiol. Behav.* **8**, 683–686 (1972).
19. Galef, G. Social identification of toxic diets by Norway rats (*Rattus norvegicus*). *J. Comp. Psychol.* **100**, 331–334 (1986).
20. Johnson, E. Pheromones, the vomeronasal system and communication: From hormonal responses to individual recognition. *Ann. N.Y. Acad. Sci.* (in the press).
21. Rottman, S. J. & Snowdon, C. T. Demonstration and analysis of an alarm pheromone in mice. *J. Comp. Physiol. Psychol.* **81**, 483–490 (1972).
22. Nixon, A., Mallet, A. I. & Gower, D. B. Simultaneous quantification of five odorous steroids (16-androstenes) in the axillary hair of men. *J. Ster. Biochem.* **29**, 505–510 (1988).
23. Cohn, B. A. In search of human skin pheromones. *Arch. Dermatol.* **130**, 1048–1051 (1994).
24. Spielman, A. I., Zeng, X. N., Leyden, J. J. & Preti, G. Proteinaceous precursors of human axillary odor: isolation of two novel odor-binding proteins. *Experientia* **51**, 40–47 (1995).
25. Russell, M. J., Switz, G. M., Thompson, K. Olfactory influences on the human menstrual cycle. *Pharmacol. Biochem. Behav.* **13**, 737–738 (1980).
26. Preti, G., Cutler, W. B., Garcia, C. R., Huggins, G. R. & Lawkey, H. J. Human axillary secretions influence women's menstrual cycles: The role of donor extract from females. *Horm. Behav.* **20**, 474–482 (1986).
27. Doty, R. Olfactory communication in humans. *Chem. Senses* **6**, 351–376 (1981).
28. Wilson, H. C. Female axillary secretions influence women's menstrual cycles: A critique. *Horm. Behav.* **21**, 536–546 (1987).
29. Stern, K. & McClintock, M. K. In *Psychopharmacology of Women* (eds Jensvold, M. F., Halbreich, U. & Hamilton, J.) 393–413 (Am. Psychiatry Press, Washington DC, 1996).
30. Doty, R. L., Snyder, P. J., Huggins, G. R. & Lowry, L. D. Endocrine, cardiovascular and psychological correlates of olfactory sensitivity changes during the human menstrual cycle. *J. Comp. Physiol. Psychol.* **95**, 45–60 (1981).

Acknowledgements. We thank T. Z. Snyder for editing and J. Altmann, J. Charrow, S. Fisher and S. Goldin-Meadow for their comments. This work was supported by the US NIMH and NIH and a grant from the John D. and Catherine T. MacArthur Foundation.

Correspondence and requests for materials should be addressed to M.K.M. (e-mail: mkm1@midway.uchicago.edu).

Motor role of human inferior parietal lobe revealed in unilateral neglect patients

Jason B. Mattingley*, Masud Husain†, Chris Rorden‡, Christopher Kennard† & Jon Driver‡

* Department of Experimental Psychology, University of Cambridge, Cambridge CB2 3EB, UK

† Division of Neuroscience and Psychological Medicine, Imperial College School of Medicine, Charing Cross Hospital, London W6 8RF, UK

‡ Institute of Cognitive Neuroscience, Department of Psychology, University College London, Gower Street, London WC1E 6BT, UK

The exact role of the parietal lobe in spatial cognition is controversial. One influential hypothesis proposes that it subserves spatial perception¹, whereas other accounts suggest that its primary role is to direct spatial movement^{2,3}. For humans, it has been suggested that these functions may be divided between inferior and superior parietal lobes, respectively^{2,4}. In apparent support of

a purely perceptual function for the inferior parietal lobe (IPL), patients with lesions to this structure, particularly in the right hemisphere, exhibit unilateral spatial neglect (deficient awareness for the side of space opposite to that of their lesion)⁵. Here we show that patients with right IPL lesions also have a specific difficulty in initiating leftward movements towards visual targets on the left side of space. This motor impairment was not found in neglect patients with frontal lesions, contrary to previous proposals that motor aspects of neglect are particularly associated with anterior damage⁶⁻⁹. Our results suggest that the human IPL operates as a sensorimotor interface, rather than subserving only perceptual functions.

Unilateral neglect patients often fail to acknowledge stimuli located contralateral to their lesion, or are slow to respond to them, despite relatively preserved afferent inputs on the affected side⁵. Although neglect clearly has a perceptual component¹⁰⁻¹³, it may also involve a directional bias in motor control, which disadvantages movements toward contralesional stimuli^{6-9,14}. Previous evidence for motor deficits in neglect has come mainly from patients with large lesions involving the frontal lobe⁷⁻⁹, rather than damage restricted to just the IPL, and from tasks which sought to isolate motor components of neglect by having patients make unnatural movements away from visual targets^{7,8}. One problem with such tasks is that patients with frontal lobe lesions may perform abnormally because of their general difficulties with highly incompatible responses^{15,16}.

Our study introduces a new method which allows separation of

sensory and motor components of neglect, without ever requiring movements away from targets. Rather than applying this test to patients with large lesions extending across both the parietal and frontal lobes, we studied two separate groups with focal lesions. The lesions of the posterior group were centred on the right IPL, allowing a critical test of whether this structure is purely perceptual. The anterior group had focal lesions centred on the right inferior frontal lobe (IFL). Damage here can also produce neglect^{17,18}, but no study has contrasted IFL with IPL lesion patients to test whether motor aspects of neglect depend on anterior damage.

In our reaching task, sensory information about target location remained constant while the direction of movement required to reach towards the target was manipulated. Visual fixation was maintained at a central position while the start position of the responding hand was varied. Visual targets could appear on the left or right of fixation. Patients made fast reaches with their right hand towards whichever light turned green. In separate blocks of trials the right hand started: midway between the potential targets (at the body midline); to the extreme left of both targets; or to the extreme right (Fig. 1a-c).

In left neglect patients, reaction times to a left target should be slower than those to a right target when the hand starts centrally (Fig. 1a). This left delay could be due to impaired perceptual attention for left targets, or to a problem in initiating leftward movements, or to some combination of these deficits. The critical situation for evaluating any directional motor impairment arose when the hand started from an extreme left position (Fig. 1b). In this case, a rightward movement was now required to reach towards a target on the left of fixation. If delayed reactions to left targets from a central start were caused by a problem in initiating leftward movements, then this delay should be reduced when only rightward movements were required, from the left start. We also examined reaches with an extreme right start, from which all movements were leftward (Fig. 1c).

Six patients with left neglect after right-hemisphere stroke were tested, at a mean of 49 days post stroke. Lesion reconstructions¹⁹ are shown in Fig. 2. The region of cortical lesion overlap for the three patients with posterior strokes was in the IPL (Fig. 2c); none of their

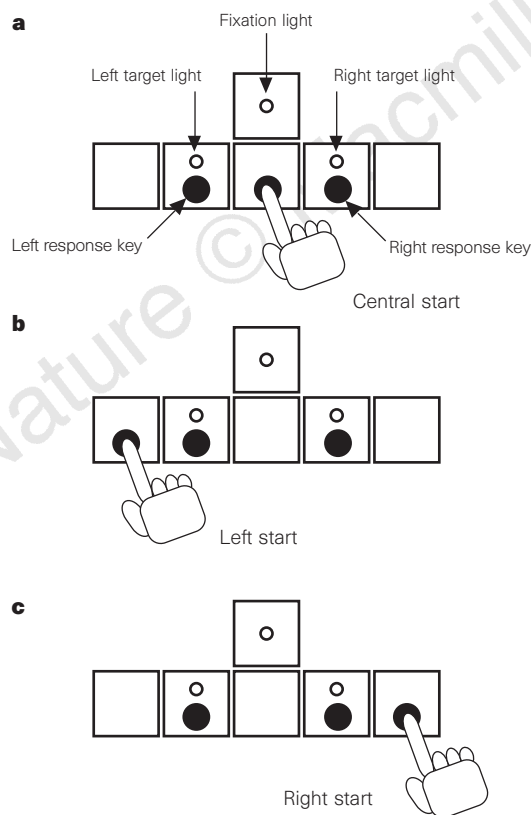


Figure 1 Reaction times to left or right visual targets were measured in a reaching task (experiment 1) and a no-reach task (experiment 2) using the same apparatus. In both conditions, the hand was positioned, in separate blocks of trials, either at a central (a), extreme left (b), or extreme right (c) start position. In the reaching task, subjects had to move their hand from the start position to press the key immediately beneath the green target. In the no-reach task, their response was to depress the start key on detecting a green target, and so no spatial reach was required.

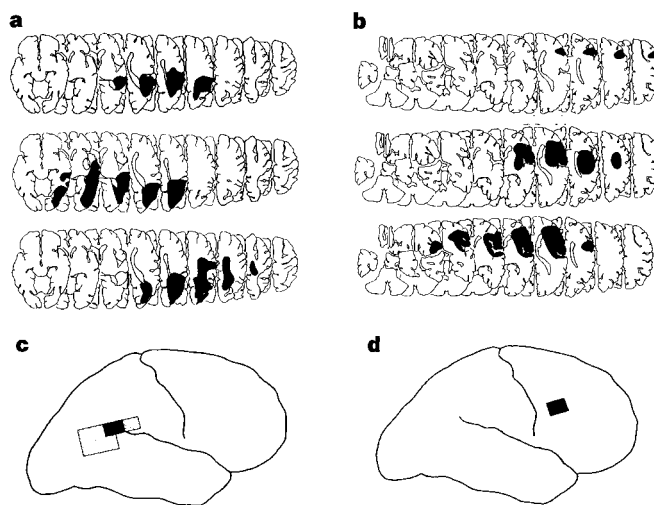


Figure 2 Extent of right-hemisphere lesions in neglect patients. a, Posterior or b, anterior strokes, plotted from CT scans on to axial templates¹⁹. Lateral views of the brain, demonstrating the common cortical (solid) and subcortical (hatched) lesion, are shown in (c) and in (d) for patients with posterior or anterior lesions, respectively. The common cortical area in the posterior group occupies a small region of the right inferior parietal lobe (IPL); for the anterior group, it occupies a small region of the right inferior frontal lobe (IFL). Note that for patients with anterior lesions, the common subcortical zone lies immediately beneath the common cortical zone.

lesions involved the superior parietal cortex. The region of cortical lesion overlap for the three anterior patients was in the IFL (Fig. 2d). All six subjects showed left neglect on standard clinical tests. On the Mesulam shape-cancellation task²⁰ the IPL patients found a mean of only 33/60 targets, all on the right of the sheet. The IFL patients found a mean of 34/60 targets (all on the right), showing comparable severity of clinical neglect. No patient had optic ataxia or apraxia.

Mean reaction times to initiate a reach for left versus right targets from each start position are shown in Fig. 3, separately for IPL and IFL groups. IPL patients showed substantially slower reaction times for left targets than right targets when the hand started centrally, indicating left neglect. In the critical condition where the hand started from the extreme left (so that left targets now required a rightward movement), the disadvantage for left targets was dramatically reduced. This pattern was found for all three IPL patients, and suggests that a substantial component of their difficulty with left targets from a central start was due to a motor deficit in initiating leftward reaches, rather than solely to a perceptual deficit in detecting left targets.

When the hand started from the extreme right, reaction times for IPL patients were as for a central start; reaches were still initiated more slowly for left than right targets, even though both required leftward movements. The motor impairment in the IPL patients is therefore not simply a difficulty in initiating any leftward movement; rather it is a specific difficulty in initiating leftward movements to targets located in the left visual field. Thus, both sensory and motor factors determine the IPL patients' performance.

For the IFL group, hand start position had no significant effects on reaction time. All three IFL patients reacted more slowly to left than right targets (revealing neglect), but did so regardless of the direction of movement required. Because the IFL group did not show the motor initiation deficit identified in the IPL patients, they serve as a control group to show that the IPL pattern is lesion-specific. We have also confirmed that a further control group of normal elderly people do not show the IPL pattern.

The IPL pattern could perhaps be attributed to the sensory attention of those patients being 'cued' towards their neglected left side, by proprioceptive and/or visual inputs from the responding hand when it rested on the left start key. Experiment 2 tested this

alternative to our motor account, using a 'no-reach' control task. We varied hand position as before, but now the patients were only required to press the start key as soon as they detected a green target on either side. Thus, they responded where their hand was already located, rather than reaching towards the target. Central fixation was again required, so the visual events were as in experiment 1. Likewise, the afferent inputs from the right hand were also exactly as before for the three different start positions. If the previous IPL pattern was due to these afferent inputs, the new task should replicate it. If instead the critical factor was the direction of reaching required for left targets, the results should change, because directional reaches were no longer required.

Mean results from the control no-reach task are shown in Fig. 4. Unlike the reaching task (compare left graphs in Fig. 4 versus Fig. 3), changing the start position of the hand now had no effect on the IPL group. Responses were still slower overall for left targets, revealing a perceptual component to these patients' neglect (delayed detection of left visual events); but critically, there was no longer an improvement for left targets when the hand was located on the left. This excludes any explanation for experiment 1 in terms of 'cuing' by afferent inputs from the right hand when at a left start. The effect of start position in experiment 1 for IPL patients must therefore have been due to the motor deficit we propose, impairing the initiation of leftward reaches to left targets.

The IFL group showed no effect of start position in either experiment. The absence of a motor initiation deficit for leftward reaches in these anterior patients (experiment 1) seems contrary to previous suggestions of a specific association between anterior lesions and motor components of neglect⁶⁻⁹. However, previous studies did not test patients with focal frontal lesions as here, but rather with larger lesions involving both parietal and frontal lobes.

The specific motor initiation deficit found in the IPL patients has parallels with recent electrophysiological findings from monkey posterior parietal cortex. Cells there were previously thought to have primarily sensory or attentional roles^{21,22}, but a recent study observed neuronal discharges tuned not only for a sensory location, but also for the movement (a saccade or reach) being planned towards it³. Human parietal cortex may contain neurons with similar response properties, computing discrepancies between current hand position and target location²³⁻²⁵. An impairment to such

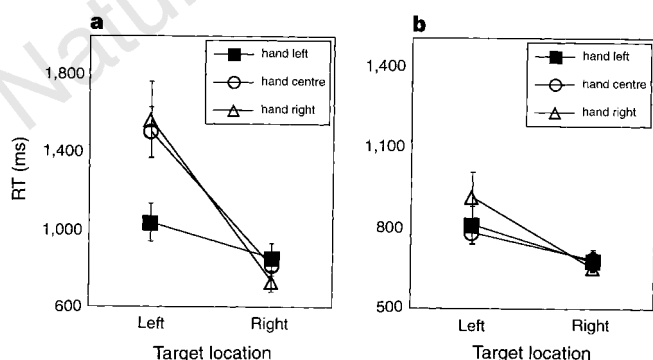


Figure 3 Mean reaction time (RT) to initiate movements towards left versus right targets, from three different start positions in the reaching task of experiment 1. **a**, IPL patients; **b**, IFL patients. With a central start, RT to left targets was slower than for right targets, consistent with left neglect ($P < 0.05$ for five patients; $P < 0.06$ for one frontal patient). For the IPL group, slower RT for left than right targets was found from the central and right start positions ($F(1) = 27.72$, $P < 0.01$ and $F(1) = 40.36$, $P < 0.01$, respectively), but critically not from the left start position ($F(1) = 1.92$, NS). In contrast, for the IFL group there was no effect of start position on RT to left versus right targets ($F(2, 4) = 2.91$, NS). Direct comparison of the two groups confirmed that the RT difference between left and right targets depended on start position for IPL patients but not IFL patients ($F(2, 4) = 8.18$, $P < 0.5$).

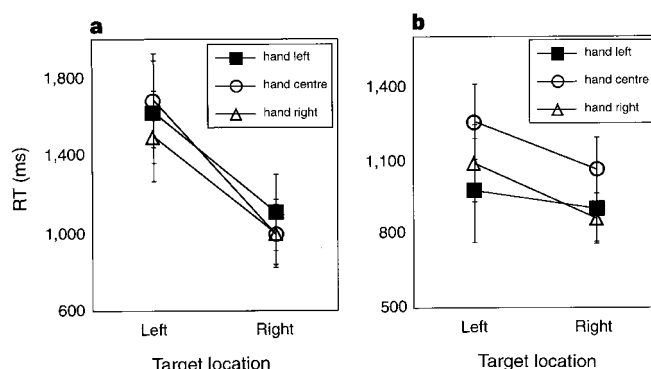


Figure 4 Mean RT to detect left versus right targets, for the three different hand positions in the no-reach task of experiment 2. **a**, IPL patients; RT was significantly longer for left than right targets, but this was unaffected by hand position ($F(2, 4) = 2.78$, NS). Unlike the reaching task of experiment 1 (compare with Fig. 3a), there was no improvement in RT to the left target when the hand was placed on the extreme left. **b**, IFL patients: RT to left versus right targets was again unaffected by hand start position ($F(2, 4) = 4.58$, NS).

parietal computations after right IPL damage might be responsible for the observed delay in initiating leftward movements to targets in the left hemispace. The parietal lobe is increasingly regarded as the final stage of a 'dorsal stream' of visual pathways (as distinct from a 'ventral' stream to the temporal lobe)¹. Milner and Goodale^{2,4} proposed that the dorsal stream directs spatial movements, consistent with the misreaching seen in patients with superior parietal lesions²⁶. But before this study, there was no evidence for motoric impairments after inferior parietal damage. Indeed, Milner and Goodale considered human IPL primarily as part of the ventral stream concerned with perceptual awareness, although they suggested it may 'co-opt' dorsal processes². Our results show for the first time that patients with right IPL lesions have an 'intentional' impairment in initiating leftward movements towards targets in the left hemispace, in addition to their perceptual difficulties. We therefore conclude that the human IPL acts as a sensorimotor interface, rather than having exclusively perceptual or motor functions. □

Methods

The apparatus comprised six white plastic boxes (each 10 cm square, 4.5 cm high) mounted on a board (Fig. 1). One box was placed centrally, further away from the patient than the remaining five, with a yellow light-emitting diode (LED) in it serving as the fixation point. Three boxes each had a microswitch connected to a response key (diameter 32 mm). Two of these boxes also contained bi-colour (green/red) LEDs which served as target (green) or distractor (red) stimuli (roughly 8° eccentricity from fixation). The response key in the third box served as the start key. This start box could be moved to occupy the central, extreme left or extreme right position (Fig. 1a–c). The remaining two boxes filled spaces unoccupied by boxes with response keys. A computer controlled the LEDs and recorded key-presses.

The patients sat with body midline aligned with the central fixation point. Eye position was monitored by an investigator, and trials with saccades before limb movement were discarded. The patients responded with the index finger of their right hand. They received at least 15 practice trials on both the reaching and no-reach tasks, being randomly assigned to start on either task. Each block of 50 experimental trials comprised a random sequence of 20 target-only trials with just a green LED illuminated (10 left; 10 right); 20 target-plus-distractor trials, with green target on one side and concurrent red distractor on the other side (10 left targets; 10 right); and 10 distractor-only ('catch') trials, with only a red distractor (5 left catch; 5 right). Response had to be withheld on catch trials. Data presented are pooled over target-only and target-plus-distractor trials, as preliminary analyses found no difference. Start position was blocked, in a randomized order across subjects.

In the reaching task of experiment 1, trials began with the onset of the yellow central fixation LED, as a signal to depress the start key. A variable 500–1,000 ms after this key was depressed, the fixation LED extinguished, and a target (green LED), a distractor (red LED), or a target on one side plus a distractor on the opposite side, were illuminated. Patients reached as quickly as possible to press the key immediately beneath the green LED. They were asked to ignore any red LED, and to withhold movements on catch trials. Reaction time to initiate the movement was recorded. On each trial, target and distractor LEDs remained illuminated until a key was pressed, or for a maximum of 4,000 ms. The 'no-reach' task used the same apparatus, but patients were instructed to depress and then gently release the start key when the fixation LED came on. When they saw a target (green LED) they had to depress the start key (on which their finger was still resting) as fast as possible. Reaching and no-reach tasks were intermingled within each session to allow a meaningful comparison.

Received 1 October; accepted 20 November 1997.

1. Ungerleider, L. G. & Mishkin, M. in *Analysis of Visual Behaviour* (eds Ingle, D. J., Goodale, M. A. & Mansfield, R. J. W.) 549–586 (MIT Press, Cambridge, MA, 1982).
2. Milner, A. D. & Goodale, M. A. *The Visual Brain in Action* (Oxford Univ. Press, Oxford, 1995).
3. Snyder, L. H., Batista, A. P. & Andersen, R. A. Coding of intention in the posterior parietal cortex. *Nature* **386**, 167–170 (1997).
4. Milner, A. D. in *Parietal Lobe Contributions to Orientation in 3D Space* (eds Thier, P. & Karnath, H.-O.) 3–22 (Springer, Heidelberg, 1997).
5. Robertson, I. H. & Marshall, J. C. (eds) *Unilateral Neglect: Clinical and Experimental Studies* (Lawrence Erlbaum, Hillsdale, 1993).

6. Mesulam, M.-M. A cortical network for directed attention and unilateral neglect. *Ann. Neurol.* **10**, 309–325 (1981).
7. Bisiach, E., Geminiani, G., Berti, A. & Rusconi, M. Perceptual and premotor factors of unilateral neglect. *Neurology* **40**, 1278–1281 (1990).
8. Tegnér, R. & Levander, M. Through a looking glass. A new technique to demonstrate directional hypokinesia in unilateral neglect. *Brain* **114**, 1943–1951 (1991).
9. Mattingley, J. B., Bradshaw, J. L. & Phillips, J. G. Impairments of movement initiation and execution in unilateral neglect—Directional hypokinesia and bradykinesia. *Brain* **115**, 1849–1874 (1992).
10. Bisiach, E. Mental representation in unilateral neglect and related disorders: the twentieth Bartlett Memorial Lecture. *Q. J. Exp. Psychol.* **46A**, 435–461 (1993).
11. Kinsbourne, M. A model for the mechanism of unilateral neglect of space. *Trans. Am. Neurol. Assoc.* **95**, 143–145 (1970).
12. Posner, M. I., Walker, J. A., Friedrich, F. J. & Rafal, R. D. Effects of parietal injury on covert orienting of attention. *J. Neurosci.* **4**, 1863–1874 (1984).
13. Riddoch, M. J. & Humphreys, G. W. The effect of cueing on unilateral visual neglect. *Neuropsychologia* **21**, 589–599 (1983).
14. Heilman, K. M., Watson, R. T. & Valenstein, E. in *Clinical Neuropsychology* (eds Heilman, K. M. & Valenstein, E.) 243–293 (Oxford Univ. Press, New York, 1985).
15. Mattingley, J. B. & Driver, J. in *Parietal Lobe Contributions to Orientation in 3D Space* (eds Thier, P. & Karnath, H.-O.) 309–338 (Springer, Heidelberg, 1997).
16. Duncan, J. in *The Cognitive Neurosciences* (ed. Gazzaniga, M. S.) 721–733 (MIT Press, Cambridge, MA, 1995).
17. Husain, M. & Kennard, C. Visual neglect associated with frontal lobe infarction. *J. Neurol.* **243**, 652–657 (1996).
18. Husain, M., Shapiro, K., Martin, J. & Kennard, C. Abnormal temporal dynamics of visual attention in patients with spatial neglect. *Nature* **385**, 154–156 (1997).
19. Damasio, H. & Damasio, A. R. *Lesion Analysis in Neuropsychology* (Oxford Univ. Press, New York, 1989).
20. Mesulam, M.-M. *Principles of Behavioural Neurology. Tests of Directed Attention and Memory* (FA Davis, Philadelphia, 1985).
21. Bushnell, M. C., Goldberg, M. E. & Robinson, D. L. Behavioural enhancement of visual responses in monkey cerebral cortex. I. Modulation in posterior parietal cortex related to selective visual attention. *J. Neurophysiol.* **46**, 755–772 (1981).
22. Duhamel, J.-R., Colby, C. L. & Goldberg, M. E. The updating of the representation of visual space in parietal cortex by intended eye movements. *Science* **255**, 90–92 (1992).
23. Andersen, R. A., Snyder, L. H., Bradley, D. C. & Xing, J. Multimodal representation of space in the posterior parietal cortex and its use in planning movements. *Annu. Rev. Neurosci.* **20**, 303–330 (1997).
24. Graziano, M. S. A., Gross, C. G. & Fernandez, T. A comparison of bimodal, visual-tactile neurons in parietal area 7b and ventral premotor cortex of the monkey brain. *Soc. Neurosci. Abstr.* **22**, 398 (1996).
25. Husain, M. in *Vision and Visual Dysfunction* Vol. 13 (ed. Stein, J. F.) 12–43 (Macmillan, Basingstoke, 1991).
26. Perenin, M.-T. & Vighetto, A. Optic ataxia: a specific disruption in visuomotor mechanisms. I. Different aspects of the deficit in reaching for objects. *Brain* **111**, 643–674 (1988).

Acknowledgements. We thank J. Wade and A. Rudd, and the staff and patients of the stroke units at Charing Cross Hospital and St. Thomas' Hospital, London. This research was supported by the Stroke Association and the Wellcome Trust. J.B.M. was supported by a National Health and Medical Research Council (Australia) Neil Hamilton Fairley Fellowship, and by AMRAD Australia.

Correspondence and requests for materials to M.H. (e-mail: m.husain@cxwms.ac.uk).

Role of the NF-ATc transcription factor in morphogenesis of cardiac valves and septum

José Luis de la Pompa*, Luika A. Timmerman†, Hiroaki Takimoto*, Hiroki Yoshida*, Andrew J. Elia*, Enrique Samper*, Julia Potter*, Andrew Wakeham*, Luc Marengere*, B. Lowell Langille‡, Gerald R. Crabtree† & Tak W. Mak*

* The Amgen Institute, Ontario Cancer Institute, and Departments of Medical Biophysics and Immunology, University of Toronto, 620 University Avenue, Toronto, Ontario M5G 2C1, Canada

† The Department of Developmental Biology, Stanford University Medical School Howard Hughes Medical Institute, 300 Pasteur Drive, Stanford, California 94305, USA

‡ Max Bell Research Centre, Toronto General Division, 200 Elizabeth Street, Toronto, Ontario M5G 2C4 and Department of Pathology, University of Toronto, Toronto, Ontario M5S 1A8, Canada

In lymphocytes, the expression of early immune response genes is regulated by NF-AT transcription factors^{1,2} which translocate to the nucleus after dephosphorylation by the Ca²⁺-dependent phosphatase, calcineurin³. We report here that mice bearing a disruption in the NF-ATc gene fail to develop normal cardiac valves and septa and die of circulatory failure before day 14.5 of

development. *NF-ATc* is first expressed in the heart at day 7.5, and is restricted to the endocardium, a specialized endothelium that gives rise to the valves and septum. Within the endocardium, specific inductive events appear to activate *NF-ATc*: it is localized to the nucleus only in endocardial cells that are adjacent to the interface with the cardiac jelly and myocardium, which are thought to give the inductive stimulus to the valve primordia⁴. Treatment of wild-type embryos with FK506, a specific calcineurin inhibitor⁵, prevents nuclear localization of *NF-ATc*. These data indicate that the Ca^{2+} /calcineurin/*NF-ATc* signalling pathway is essential for normal cardiac valve and septum morphogenesis; hence, *NF-ATc* and its regulatory pathways are candidates for genetic defects underlying congenital human heart disease.

Phosphorylation of amino-terminal serine residues controls transcriptional activity of NF-AT family members by influencing their nuclear localization⁶. In *NF-ATc* these amino acids are encoded by exon 3, and we deleted this by homologous recombination in embryonic stem (ES) cells (Fig. 1a, b). No homozygous mutant offspring was identified among 91 pups from heterozygous intercrosses. At E12.5, 50% of the *NF-ATc*^{-/-} mutant embryos were anaemic, oedematous and showed signs of abdominal necrosis (Fig. 1c), and died by E14.5.

Mutant embryos do not express any *NF-ATc* transcript (Fig. 1d)

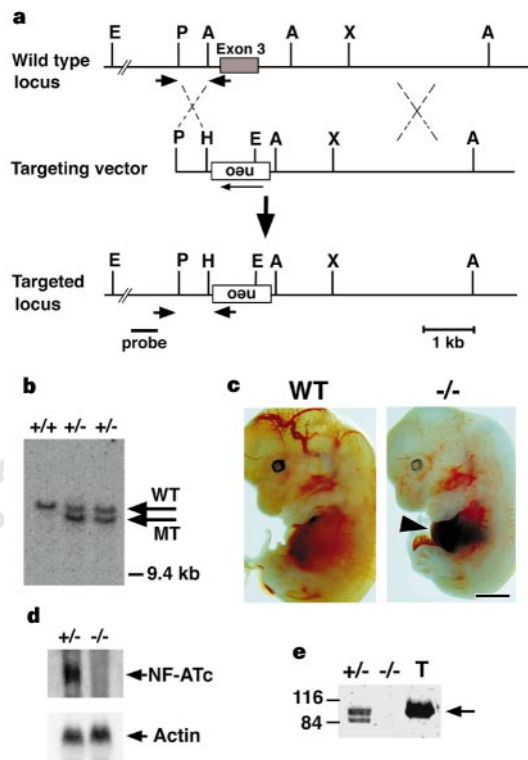


Figure 1 Targeting of the *NF-ATc* gene. **a**, Top: structure of the genomic DNA encoding exon 3 of *NF-ATc*. Arrows depict PCR primers used to identify the wild-type allele. Middle: targeting vector designed to replace exon 3 by a *neo* cassette. Bottom: targeted locus. Arrows depict PCR primers used to identify the mutant allele. A, *Apal*; E, *EcoRI*; H, *HindIII*; P, *PstI*; X, *XbaI*. Bar, 5' flanking probe. **b**, Southern blot of *EcoRI*-digested genomic DNA from wild-type (+/+) and heterozygous (+/-) ES cell lines hybridized with a 5' flanking probe. The wild-type (wt) and mutant (mt) alleles bands are indicated by arrows. **c**, E12.5 wild-type (left) and *NF-ATc*^{-/-} mutant (right) embryos. The arrowhead points to the necrotic abdominal region in the mutant embryo. Scale bar, 500 μm . **d**, Northern blot of RNA isolated from *NF-ATc*^{+/-} and *NF-ATc*^{-/-} E11.5 embryos, hybridized with a 3' *NF-ATc* probe. **e**, Western blot performed with the MAb 7A6 (ref. 20) on E11.5 *NF-ATc*^{+/-}, *NF-ATc*^{-/-} mutant embryos and mature T cells (T).

and lack the epitope encoded by exon 3 (Fig. 1e). During embryogenesis, *NF-ATc* messenger RNA was restricted to the heart (Fig. 2), and was initially found in the endocardial tubes of the E7.5 embryo (Fig. 2a), and in the endocardium and vitelline vein by E8.5 (Fig. 2b). Between E9 and E11.5, expression was observed in the endocardium of the ventricles, atrium, outflow tract (Fig. 2c–f, h) and the apex of the ventricular septum (Fig. 2h). During E11.5–E13.5, *NF-ATc* was found in the lining of the pulmonary (Fig. 2i), aortic (Fig. 2j) and atrioventricular (AV) valves (Fig. 2k) and cushion (Fig. 2l). Sections of E11.5 *NF-ATc*^{-/-} mutant embryos did not show any detectable signal (Fig. 2m), indicating that the *NF-ATc*^{-/-} is a null mutation.

To assess whether *NF-ATc* was transcriptionally active in endocardial cells, its subcellular localization was investigated by immunofluorescence. At E10.5, nuclear *NF-ATc* delineated the endocardium of the heart, that was also obtained by the endothelial marker PECAM-1 (CD31)⁷ (Fig. 2c–c). Nuclear *NF-ATc* is found primarily in cells lining the endocardial cushions, the valve and septal primordia, and along the lining of the bulbus cordis (data not shown and Fig. 3d, e). Cells exhibiting nuclear versus cytoplasmic staining often lay adjacent to one another (Fig. 3e), suggesting that short-range signals may induce *NF-ATc* activation in specific endocardial cells. No protein is detected in mesenchymal cells

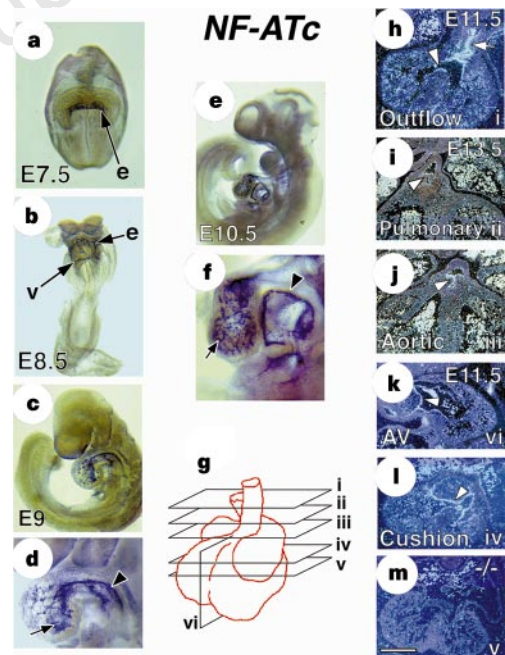


Figure 2 *NF-ATc* expression is restricted to the endocardium, outflow tract and cardiac valves. **a–f**, Whole-mount *in situ* hybridization. **a**, Frontal view of E7.5 wild-type embryo. *NF-ATc* signal (purple) is restricted to the endocardial tubes (e). **b**, E8.5 embryo. *NF-ATc* is in the endocardium (e) and vitelline vein (v). **c, d**, E9.0 embryo. *NF-ATc* is in the endocardium of the ventricle (arrow in **d**) and outflow tract (arrowhead). **e, f**, E10.5 embryo. *NF-ATc* mRNA detected in the ventricle (arrow in **f**) and atrium (arrowhead). Embryos were hybridized with a *NF-ATc* exon 3 antisense probe. **g**, Cartoon showing the approximate plane of section indicated by Roman numbers in the bottom right of **h–m**. Radioactive *in situ* hybridization showing *NF-ATc* expression in E11.5 and E13.5 wild-type (**h–l**) and in E11.5 *NF-ATc*^{-/-} mutant embryos (**m**) in dark field views. **h**, E11.5. *NF-ATc* is in the outflow tract (arrow) and lining the apex of the ventricular septum (arrowhead). **i, j**, E13.5. *NF-ATc* localizes to the developing pulmonary and aortic valves (arrowheads). **k**, E11.5. *NF-ATc* expression in the AV valves endocardium (arrowhead). **l**, *NF-ATc* expression in the endocardial lining of the cushion (arrowhead). **m**, E11.5. *NF-ATc*^{-/-} mutant embryo. No *NF-ATc* signal is detected. Sections were hybridized with a *NF-ATc* 3' antisense probe. The bright cells in the atrial and ventricular cavities are unlabelled refracting red blood cells. Bar, 400 μm in **a**, 500 μm in **b**, 600 μm in **c**, 200 μm in **d**, 1.3 mm in **e**, 300 μm in **f**, 250 μm in **h–m**.

within the cushion, indicating that cells undergoing the epithelial to mesenchymal transition leading to cushion formation⁴ downregulate NF-ATc expression. By E11.5, nuclear NF-ATc becomes restricted to the lining of the outflow tract (Fig. 3f, g) and developing valves (Fig. 3h). At E12.5, only cells which line the immediate valve surfaces (Fig. 3i) and the ends of the expanding septa (not shown) exhibited nuclear NF-ATc staining. No staining was found in sections through adult and newborn heart (data not shown), indicating that NF-ATc activity is selectively required during embryonic cardiac development.

NF-ATc³ mutant embryos showed cardiac abnormalities by E12.5. At this time, the outflow tract in the wild-type embryo is divided into the aortic sac and pulmonary trunk, and the ventricular chambers show a complex pattern of myocardial trabeculation (Fig. 4a). In 40% of the NF-ATc³ mutant embryos, the ventricular walls were hypertrophied and the chambers were small (Fig. 4b). In these embryos, the proximal region of the outflow tract (bulbus) was severely narrowed or occluded (Fig. 4b, arrowhead), and the aortic sac had thick walls and displayed a narrow lumen that disappeared as the sac approached the conus arteriosus (data not shown). By E13.5, the NF-ATc³ mutant embryos displayed abnormalities of valve structure (Fig. 4d–f). Although septation into right ventricular (pulmonary) and left ventricular (aortic) outflow tract was normal, both semilunar valves were underdeveloped, presenting the poorly organized appearance of earlier stages (Fig. 4d, f). The tricuspid and mitral valves were also primitive in comparison to wild type and displayed no well defined leaflets

(compare Fig. 4g with h). In addition, an interventricular foramen was still present in most NF-ATc³ mutant embryos (Fig. 4i), indicating defective septum formation. Mice generated from the two independently targeted ES cells showed similar phenotypes.

Vascular endothelium, myocardium and extracellular matrix appeared unaffected in NF-ATc³ mutant embryos, as indicated by the normal expression of PECAM-1 (Fig. 4j), *cardiac-actin* (Fig. 4k) and fibronectin (data not shown). Although the neural crest participates in the formation of the pulmonary and aortic valves⁸, NF-ATc is exclusively expressed in the endocardium; thus it is unlikely that defective neural crest differentiation is involved in the abnormal valve development of NF-ATc³ mutant embryos.

Death of NF-ATc³ mutant mice is probably caused by severe defects in forward blood flow resulting from valvular incompetence, stenosis and conotruncal abnormalities. The peripheral oedema and ventricular hypertrophy may be secondary to congestive heart failure and valvular malfunction.

Morphogenesis of the endocardial cushion is induced by sustained

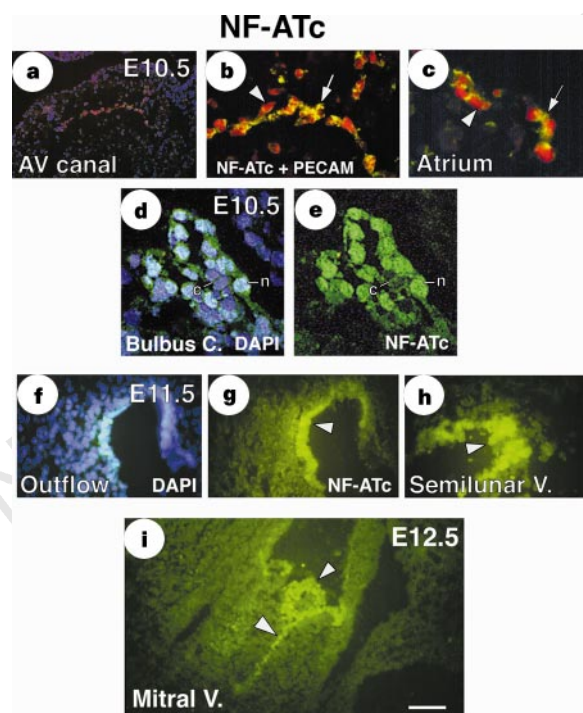


Figure 3 Active NF-ATc is expressed in a subset of endothelial cells lining the endocardial cushions and valves. Wild-type E10.5–E12.5 sections stained with anti-NF-ATc and anti-PECAM-1 (a–c), or anti-NF-ATc alone (d–i). a, b, f, DAPI counterstaining. a, General view of the atrioventricular (AV) canal region at E10.5. Note the complementary nuclear NF-ATc (red fluorescence, arrowhead) and cytoplasmic PECAM-1 stainings (arrow) in the AV canal (b) and lumen of the atrium (c). PECAM-1 staining appears yellow because of the overlap of the green and red fluorescence of minority cytoplasmic NF-ATc staining. E10.5. Section through the base of the bulbus cordis (d), showing NF-ATc nuclear signal (e). C, cytoplasmic; n, nuclear. f, g, Nuclear localization of NF-ATc in the outflow tract (g, arrowhead) and semilunar valve (h, arrowhead) of E11.5 embryos. i, E12.5 embryo showing nuclear NF-ATc in the AV canal (large arrowhead) and in the forming mitral valve (arrowhead). Scale bar, 63 μ m in a, 25 μ m in b, d, e–h, 16 μ m in c.

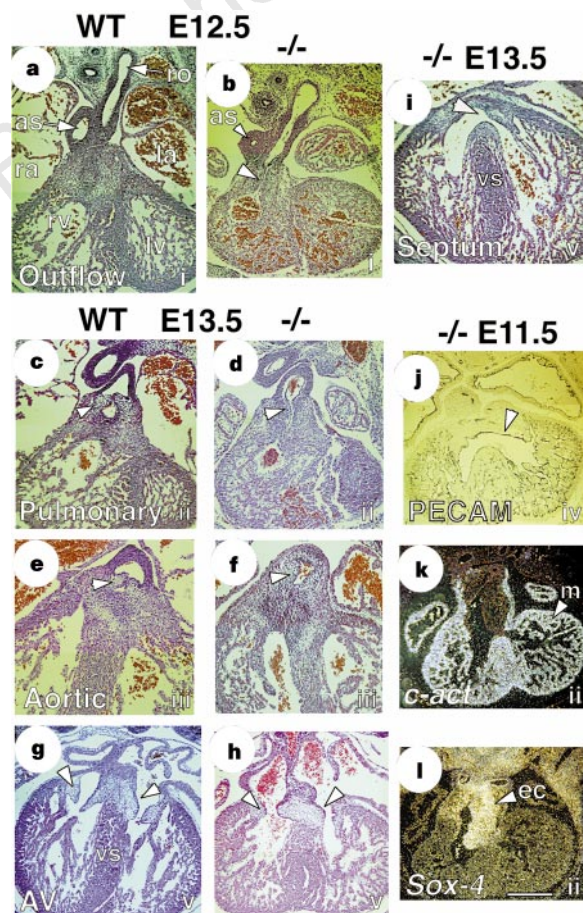


Figure 4 Cardiac valves and ventricular septum defects in NF-ATc³ mutant embryos. H + E stainings of transversal sections from wild-type (a, c, e, g) and NF-ATc³ mutant (b, d, f, h) embryos. Roman numbers (bottom right) indicate the approximate plane of section according to Fig. 2g. Section through the outflow tract (a, b) showing the stenosis in E12.5 mutant embryos (arrowhead in b) and the reduced lumen of the aortic sac (as). E13.5 embryos sectioned through the pulmonary (c, d, arrowheads), aortic (e, f, arrowheads) and atrioventricular (AV) (g, h, arrowheads) valves. i, E13.5 NF-ATc³ mutant. A septum foramen is present (arrowhead). j–l, E11.5 NF-ATc³ mutant embryos. Immunostaining showing PECAM-1 expression lining the endocardium (j, arrowhead). Radioactive *in situ* hybridization showing *c-act* expression (k) in the myocardium (m) and Sox-4 expression (l) in endocardial cushion (ec). Abbreviations: la, left atrium; lv, left ventricle; ra, right atrium; rv, right ventricle; ro, right ventricular outflow; vs, ventricular septum. Scale bar, 250 μ m.

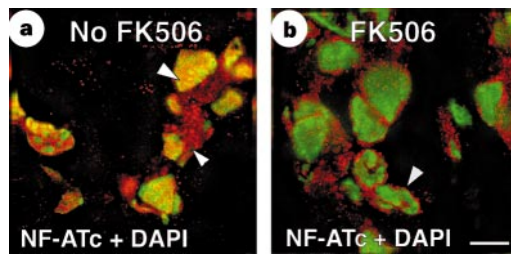


Figure 5 FK506 treatment blocks NF-ATc nuclear import in the endocardium. **a**, NF-ATc (red) and DAPI (green) stainings in endocardial cells from a wild-type control embryo cultured for 1 h with ethanol carrier. Overlap of NF-ATc and DAPI stainings (orange) demonstrates nuclear NF-ATc localization (large arrowhead), coexisting with cytoplasmic (red) distribution (small arrowhead). **b**, FK506 treatment prevents nuclear localization of NF-ATc (red), as revealed by red cytoplasmic signal in the periphery of the cells (arrowhead). Scale bar, 6 μ m.

signals from the AV myocardium that cause cell migration and differentiation into cushion tissue mesenchyme⁴. These events can be mimicked by calcium ionophores and activators of protein kinase C⁹. To determine if the nuclear localization of NF-ATc in the embryonic endocardium is due to Ca^{2+} /calcineurin signalling, we cultured embryos in the presence or absence of FK506. Control embryos showed NF-ATc nuclear localization selectively in those cells that lie adjacent to the myocardium and cardiac jelly (Fig. 5a). In contrast, embryos treated with FK506 showed cytoplasmic localization of NF-ATc in all endocardial cells (Fig. 5b). These results suggest that the nuclear localization of NF-ATc in the endocardium is related to the local activation of calcineurin, and implicate the Ca^{2+} /calcineurin signalling pathway in cardiac valve and septum development. In this pathway, the upstream activator of NF-ATc is almost certainly calcineurin. However, the ligand and receptor that lead to NF-ATc nuclear localization, and its downstream targets, are unknown. The expression of *Sox-4*, a member of the Sry family of transcription factors¹⁰ that is required for cardiac valve development and lymphoid activation¹¹ was normal in NF-ATc³ mutant embryos (Fig. 4l). This indicates that *Sox-4* is not likely to be a target for NF-ATc, although both genes may have converging roles in heart and lymphocyte development.

An intriguing feature of Ca^{2+} signalling is that although Ca^{2+} regulation is ubiquitous, both Ca^{2+} and calcineurin have been implicated in some of the most specific biological responses including axonal guidance¹², memory¹³ and lymphocyte activation by antigens^{3,14,15}. Our results suggest that at least one way by which ubiquitous calcium signals could direct specific biological processes is through the tissue-specific expression of target molecules such as NF-ATc, involved in cardiac valve formation in the embryo and activation of immune response. Furthermore, as NF-AT family members are expressed in different tissues¹⁶, specificity of response to a Ca^{2+} stimulus could be refined by both different tissue distribution and DNA sequence recognition¹⁷.

Ventricular septal and valve defects are the most frequent of human congenital cardiac defects, present in nearly 1% of all births¹⁸. Future analysis will shed light on the potential involvement of NF-ATc and its regulatory pathways^{1,19} in human cardiac abnormalities. □

Methods

Generation of NF-ATc³ mutant mice. A targeting vector was designed to replace a 1.5 kb genomic fragment containing exon 3 with the pMC1neo resistance expression cassette in reverse orientation to NF-ATc transcription. The targeting vector was linearized and electroporated into E14K ES cells. After G418 selection, homologous recombinants were identified by polymerase chain reaction (PCR) and confirmed by Southern blot. PCR, Southern, northern and western blots were performed according to standard procedures. The anti-NF-ATc monoclonal antibody 7A6 (ref. 20) recognizes the epitope deleted by the

NF-ATc³ mutation (amino acids 176–380). The mutant phenotype was analysed in C57BL/6 and CD1; no differences were observed.

Histological analysis. Embryos were fixed in 4% paraformaldehyde, dehydrated, embedded in wax, sectioned and processed for H + E staining according to standard protocols.

Whole-mount *in situ* hybridization. Embryos were fixed and processed following published protocols²¹. The NF-ATc probes used were a 1.6 kb 5' NF-ATc cDNA fragment containing exon 3, and a 900 bp 3' NF-ATc cDNA fragment, downstream to the region deleted by the NF-ATc³ mutation.

Radioactive *in situ* hybridization. Embryos were isolated and processed as for histological analysis. Hybridization was performed following published protocols²². The NF-ATc probes used were the ones cited above; the α -cardiac actin probe was a 130 bp 5' fragment²³; the *Sox-4* probe was a 300 bp 5' fragment¹¹.

Immunohistochemistry. Wild-type and mutant embryos were isolated and processed for histology as above. Staining was performed using a rat antibody against PECAM-1 (Pharmingen, 1:200 dilution), a biotinylated secondary antibody against rat IgG, and avidin-conjugated peroxidase (Vector), according to published protocols²⁴.

Immunofluorescence. Embryos were dissected out of the yolk sac and immediately embedded in OCT (Tissue-Tek), frozen and sectioned. The anti-NF-ATc antibody 7A6 was used at 1:500. The signal was enhanced by an amplification step with anti-mouse IgG-biotin (CALTAG) 1:500, and avidin-Rhodamine (Vector) 1:2,000 (Figs 3a–c and 5), or avidin D-FITC (Vector) 1:1,000 (Fig. 3e, g, h, l). The anti-PECAM-1 rat antibody (Pharmingen) was used at 1:200 followed by incubation with an anti-rat IgG-FITC (CALTAG) at 1:400. Nuclei were counterstained with DAPI (Molecular Probes, 2.5 μ g ml⁻¹) in PBS.

Short-term embryo culture. Embryos were collected at day E8.5 from plugged CD1 females, dissected out of Reichert's membrane, and cultured at 37 °C in DMEM containing 25% fetal bovine serum and 25% rat serum, using a BTC rotating bottle culture unit with continual gassing with 5% CO₂ in balanced air. Embryos were cultured for 1 h in the presence or absence of 50 nM FK506, dissected away from the yolk sacs, and embedded in OCT for cryosectioning and staining with 7A6.

Received 17 September; accepted 2 December 1997.

- Shaw, J. P. *et al.* Identification of a putative regulator of early T cell activation genes. *Science* **241**, 202–205 (1988).
- Flanagan, W. M., Corthesy, B., Bram, R. J. & Crabtree, G. R. Nuclear association of a T-cell transcription factor blocked by FK-506 and cyclosporin A. *Nature* **352**, 803–807 (1991).
- Cliststone, N. A. & Crabtree, G. R. Identification of calcineurin as a key signalling enzyme in T-lymphocyte activation. *Nature* **357**, 695–697 (1992).
- Eisenberg, L. M. & Markwald, R. R. Molecular regulation of atrioventricular valvuloseptal morphogenesis. *Circ. Res.* **77**, 1–6 (1995).
- Liu, J. *et al.* Calcineurin is a common target of cyclophilin-cyclosporin A and FKBP-FK506 complexes. *Cell* **66**, 807–815 (1991).
- Beals, C. R., Cliststone, N. A., Ho, S. N. & Crabtree, G. R. Nuclear localization of NF-ATc by a calcineurin-dependent, cyclosporin-sensitive intramolecular interaction. *Genes Dev.* **11**, 824–834 (1997).
- Buck, C. A. *et al.* Cell adhesion receptors and early mammalian heart development: an overview. *C. R. Acad. Sci. III* **316**, 838–859 (1993).
- Kirby, M. L., Gale, T. F. & Stewart, D. E. Neural crest cells contribute to normal aorticopulmonary septation. *Science* **220**, 1059–1061 (1983).
- Runyan, R. B. *et al.* Signal transduction of a tissue interaction during embryonic heart development. *Cell Regul.* **1**, 301–313 (1990).
- Gubbay, J. *et al.* A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. *Nature* **346**, 245–250 (1990).
- Schilham, M. W. *et al.* Defects in cardiac outflow tract formation and pro-B-lymphocyte expansion in mice lacking *Sox-4*. *Nature* **380**, 711–714 (1996).
- Chang, H. Y. *et al.* Asymmetric retraction of growth cone filopodia following focal inactivation of calcineurin. *Nature* **376**, 686–690 (1995).
- Bennett, P. C., Zhao, W., Lawen, A. & Ng, K. T. Cyclosporin A, an inhibitor of calcineurin, impairs memory formation in day-old chicks. *Brain Res.* **740**, 107–117 (1996).
- Jain, J., McCaffrey, P. G., Valge-Archer, V. E. & Rao, A. Nuclear factor of activated T cells contains Fos and Jun. *Nature* **356**, 801–804 (1992).
- O'Keefe, S. J., Tamura, J., Kincaid, R. L., Tocci, M. J. & O'Neill, E. A. FK-506- and CsA-sensitive activation of the interleukin-2 promoter by calcineurin. *Nature* **357**, 692–694 (1992).
- Hoey, T., Sun, Y. L., Williamson, K. & Xu, X. Isolation of two new members of the NF-AT gene family and functional characterization of the NF-AT proteins. *Immunity* **2**, 461–472 (1995).
- Ho, S. N. *et al.* NFATc3, a lymphoid-specific NFATc family member that is calcium-regulated and exhibits distinct DNA binding specificity. *J. Biol. Chem.* **270**, 19898–19907 (1995).
- Hoffman, J. I. Incidence of congenital heart disease: II. prenatal incidence. *Pediatr. Cardiol.* **16**, 155–165 (1995).
- Beals, C. R., Sheridan, C. M., Gardner, P. & Crabtree, G. R. Nuclear export of NF-ATc enhanced by glycogen synthase kinase-3. *Science* **275**, 1930–1934 (1997).
- Northrop, J. P. *et al.* NF-AT components define a family of transcription factors targeted in T-cell activation. *Nature* **369**, 497–502 (1994).
- Koop, K. E., MacDonald, L. M. & Lobe, C. G. Transcripts of Grg4, a murine groucho-related gene, are

- detected in adjacent tissues to other murine neurogenic gene homologues during embryonic development. *Mech. Dev.* 59, 73–87 (1996).
22. Hui, C. C. & Joyner, A. L. A mouse model of greig cephalopolysyndactyly syndrome: the extra-toes mutation contains an intragenic deletion of the *Gli3* gene. *Nature Genet.* 3, 241–246 (1993).
 23. Sassoon, D. A., Garner, I. & Buckingham, M. Transcripts of alpha-cardiac and alpha-skeletal actins are early markers for myogenesis in the mouse embryo. *Development* 104, 155–164 (1988).
 24. Trumpp, A., Blundell, P. A., de la Pompa, J. L. & Zeller, R. The chicken limb deformity gene encodes nuclear proteins expressed in specific cell types during morphogenesis. *Genes Dev.* 6, 14–28 (1992).

Acknowledgements. L.A.T., H.T. and H.Y. contributed equally to this work. We thank C. Sirard, J. Cross, V. Stambolic and R. Hakem for critical reading of this manuscript, and H. Clever for the *Sox-4* probe.

Correspondence and requests for materials should be addressed to T.W.M. (e-mail: t.mak@oci.utoronto.ca).

The transcription factor NF-ATc is essential for cardiac valve formation

Ann M. Ranger*, Michael J. Grusby*, Martin R. Hodge†‡, Ellen M. Gravalles*, Fabienne Charles de la Brousse‡, Tim Hoey‡, Craig Mickanin§, H. Scott Baldwin§ & Laurie H. Glimcher*

* Department of Cancer Biology, Harvard School of Public Health and Department of Medicine, Harvard Medical School, Boston, Massachusetts 02115, USA

‡ Tularik Inc, South San Francisco, California 94080, USA

§ Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19104, USA

Nuclear factor of activated T cells (NF-AT) is the name of a family of four related transcription factors that may be needed for cytokine gene expression in activated lymphocytes^{1–4}. Here we report that mice with a targeted disruption of the NF-ATc gene show an unexpected and dramatic defect in cardiac morphogenesis, with selective absence of the aortic and pulmonary valves, leading to death *in utero* from congestive heart failure at days 13.5–17.5 of gestation. In contrast, tricuspid and mitral valve morphogenesis is normal. NF-ATc is the first transcription factor known to be expressed only in the endothelial cells of the heart. As in T cells, nuclear translocation of NF-ATc in cardiac endothelial cells is controlled by the calcium-regulated phosphatase calcineurin^{5,6}: NF-ATc remains cytoplasmic in normal embryos cultured with cyclosporin A, an inhibitor of calcineurin. Abnormal development of the cardiac valves and septae is the most frequent form of birth defect, yet few molecular regulators of valve formation are known. Our results indicate that NF-ATc may play a critical role in signal-transduction processes required for normal cardiac valve formation.

To determine whether the NF-ATc protein had unique functions *in vivo*, the NF-ATc gene was disrupted in embryonic stem (ES) cells by replacing 19 amino acids (635–653) of the Rel-homology domain with a neomycin-resistance gene (*neo*). This resulted in a frameshift of the remaining amino acids of the protein (654–710) (Fig. 1a). Two of six ES clones, named 141 and 112, transmitted the disrupted allele to offspring and NF-ATc^{+/-} mice were intercrossed to generate NF-ATc^{-/-} mice. Of 181 pups born, there were no homozygotes and a slightly lower number of heterozygotes than expected, suggesting that NF-ATc is required for survival. Serial timed matings produced the expected mendelian ratios of +/+, +/- and -/- animals at embryonic day of development (e) 8.5–12.5 (Table 1). From e13.5–17.5, however, increasingly fewer live -/- animals were present, and at birth no surviving -/- embryos were found. Figure 1b shows a Southern blot analysis of e14.5 embryos representative of the three genotypes.

By e14.5, the -/- embryos sometimes appeared pale and oedematous, with frequent pericardial effusions and abdominal haemor-

rhages. The fact that these animals died at e13.5–17.5 suggested a primary cardiac or haematopoietic defect⁷. A consistent pattern of defects was seen in embryos at e14.5 (Fig. 2) and e13.5 (not shown). There was a conspicuous absence of both the pulmonary and the aortic valve leaflets, but there was normal partitioning of the large vessels into a clearly defined right ventricular, or pulmonary, outflow and left ventricular, or aortic, outflow. There was also a defect in the superior margin of the interventricular septum which resulted in a ventricular septal defect. There was no impairment in formation of the compact layer of the ventricular myocardium, or in formation of ventricular trabeculae epicardial morphogenesis, or coronary vascularization (Fig. 2). Importantly, there were no definitive abnormalities in tricuspid or mitral-valve formation or in the inferior portion of the ventricular septum, although these abnormalities are often seen with atrioventricular valve abnormalities (Fig. 3, a, b). The phenotypes of the 112 and 141 lines were qualitatively similar, although the 141 line was more severely affected: 141^{-/-} embryos were completely impaired in aortic and pulmonic valve development, whereas 112^{-/-} embryos had less pronounced valvular defects.

Valve formation is initiated when endothelial cells within the outflow tract and atrioventricular junction of the heart assume a mesenchymal phenotype and migrate into regional swellings of the extracellular matrix that are called the endocardial cushions⁸.

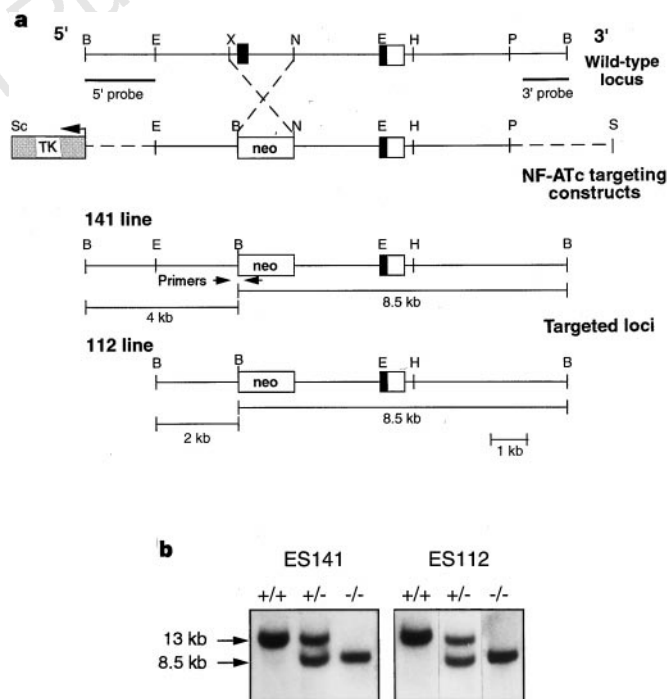


Figure 1 Targeted disruption of the murine NF-ATc gene. **a**, At the top is shown the intron-exon (exons are boxed) structure of the 3' terminus of genomic DNA encoding a portion of the NF-ATc Rel-homology domain (RHD) (black box), isolated from a 129/sv genomic library. In the middle a replacement construct containing the neo cassette, indicating the replacement of the RHD (amino acids 635–653) with neo, is shown. The region predicted to undergo replacement is indicated with dotted lines between the wild-type locus and the targeting construct. The recombinant loci for the independently derived ES clones 141 and 112 are shown at the bottom. The probes and restriction sites used for diagnostic Southern blotting are indicated (B, *Bam*H1; E, *Eco*R1; H, *Hind*III; S, *Sal*I; Sc, *Sac*I; N, *Nde*I). **b**, Analysis of DNA from e14.5 embryonic tissue of mice resulting from matings of NF-ATc^{+/-} mice from line 141 or 112. As predicted from the restriction map of the wild-type locus, digestion of DNA with *Bam*H1 generates a 13-kb fragment which is replaced by an 8.5-kb fragment when hybridized with the 3' probe. Hybridization with the 5' probe yielded the expected 13-kb band in the wild-type and 4-kb band in the 141 line but a 2-kb band in the 112 line, consistent with an additional 2-kb intronic deletion (not shown).

† Present address: Millennium Pharmaceuticals, Cambridge, Massachusetts 02139, USA

During this process, endothelial cells downregulate expression of the endothelial-cell-specific marker platelet/endothelial cell adhesion molecule-1 (PECAM-1 or CD31) (ref. 9). We studied the histology of the outflow-tract endocardial cushions (conotruncal ridges), which are the anlage of the aortic and pulmonary valves, of the e11.5 embryos to determine whether this process occurred (Fig. 3c–f). There were no qualitative differences between wild-type and $-/-$ embryos in endothelial transformation, migration, or proliferation at this critical developmental stage. Moreover, there was normal attenuation of endothelial cell expression of PECAM-1 and initiation of expression of $\alpha 4$ protein in mesenchymal cells in the mutant embryos (results not shown). These results are consistent with normal endothelial activation and transformation into mesenchyme.

Whole-mount immunohistochemistry and immunohistochemical analysis of serially sectioned embryos showed no expression of NF-ATc outside the heart, except in the thymus (results not shown). To define the temporal and tissue-specific distribution of NF-ATc during cardiovascular development, monoclonal antibodies against PECAM-1 and NF-ATc were used for double immunofluorescent labelling of wild-type and mutant embryos during critical periods of valve formation. Co-immunolocalization of PECAM-1 and NF-ATc showed conclusively that NF-ATc expression is restricted to the endothelium of the heart or endocardium (Fig. 4a–c). NF-ATc was

expressed at high levels in the endocardium of the outflow tract and atrioventricular canal region and at much lower levels in the endocardium lining the ventricular trabeculae. No expression of NF-ATc was detected in those endothelial cells that had transformed into mesenchyme and begun invasion of the extracellular matrix (Fig. 4a, b), indicating that NF-ATc may not be involved in maintenance of the mesenchymal phenotype. Furthermore, although PECAM expression is limited to the cell membrane (as seen by green fluorescence), NF-ATc expression is clearly nuclear (as shown by red fluorescence), suggesting that NF-ATc is in its active, dephosphorylated form. NF-ATc was not detected in endothelium outside the heart, as shown by the lack of fluorescence in the pharyngeal arch arteries, dorsal aorta and liver. There was no detectable NF-ATc protein in e11.5 141 and 112 $-/-$ embryos (Fig. 4c), demonstrating a null mutation of NF-ATc in these embryos.

The timing of expression of NF-ATc in the heart was also very restricted. Reverse transcription with polymerase chain reaction (RT-PCR) using NF-ATc-specific primers yielded products in the e7.5 embryo (the period of initial endocardial and myocardial differentiation^{9,10}), and at e8.5 and e12.5 (not shown). Immunohistochemistry showed prominent intranuclear NF-ATc expression that was restricted to the straight heart tube of the e8.5 embryo (Fig. 4f, g), suggesting an active role for NF-ATc at least 24 h before the

Table 1 Embryonic lethality of NF-ATc-null mice

Day	ES 141				ES 112			
	Number of mice			Viable (%)	Number of mice			Viable (%)
	+/+	+/-	-/-	-/-	+/+	+/-	-/-	-/-
e9.5	2	10	1	100	2	5	1	100
e10.5	16	26 (4)	7 (2)	71	8 (1)	26	7	100
e11.5	12	20	23 (1)	96	23	31 (2)	16	100
e12.5	4	11	4	100	12	19	12 (3)	75
e13.5	20	27 (2)	13 (2)	85	19	39 (2)	13 (1)	92
e14.5	20	26 (3)	17 (5)	71	20	48	19 (6)	68
e15.5	2	8	2 (1)	50	6	19	4 (3)	25
Postnatal	37	46	0		39	59	0	

The numbers in parentheses refer to the number of embryos that were determined to be dead or dying upon visual inspection (resorbing or no heartbeat). These numbers are included in the totals for the $+/-$ or $-/-$ embryos. There were many resorptions observed after e10.5 which could not be genotyped.

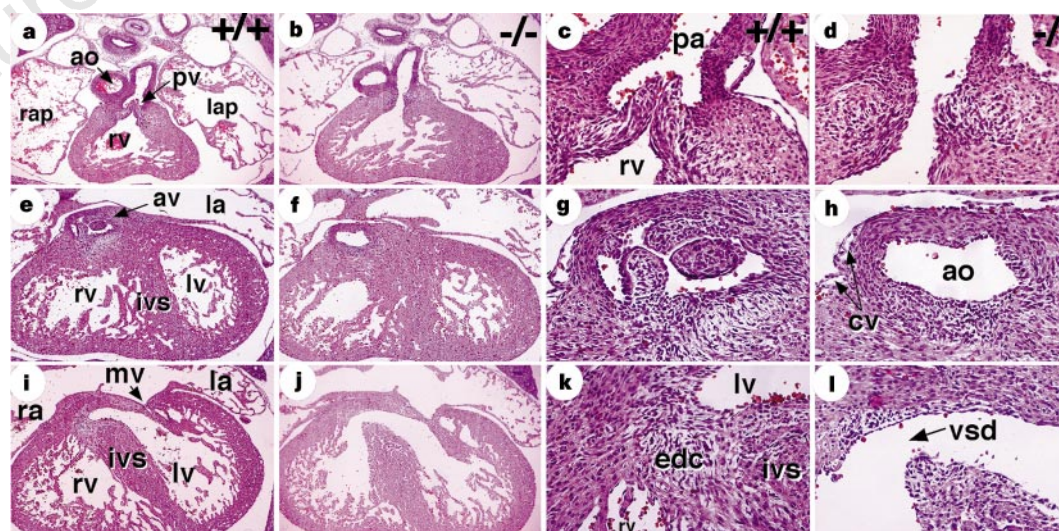


Figure 2 Histological analysis of e14.5 wild-type and NF-ATc-mutant embryos from line 141. **a, b**, Show low-power magnification of comparable cross-sections through the right ventricular outflow tracts (rv) of wild-type (**a**) and mutant (**b**) embryos. **c, d**, Show high-power magnification of null-mutant embryos (**d**) and controls (**c**). **e, f**, Show low-power magnification of comparable cross-sections through the left ventricular outflow tract (lv) of wild-type (**e**) and null-mutant (**f**) embryos. High-power magnifications revealed severe defects in formation of the

aortic valve leaflets (av) in the null $-/-$ mutant embryos (**h**) when compared with controls (**g**). **j, l**, Show the ventricular septal defect (vsd) in the interventricular septum (ivs) of NF-ATc-null mutants and **i, k** show the situation in wild-type littermates. Ao, aorta; pv, pulmonary vein; aps, aortico-pulmonary septum; pa, pulmonary artery; cv, coronary vessels; ra, right atrium; rap, right atrial appendage; la, left atrium; lap, left atrial appendage; mv, mitral valve; edc, endocardial cushion.

first evidence of valve formation. This burst of NF-ATc expression was temporary, as, by e13.5, NF-ATc protein could not be detected by RT-PCR or by immunohistochemistry (not shown). These results support a unique temporally and tissue-restricted role for NF-ATc in the signal-transduction processes required for normal cardiac valve formation and ventricular septation.

The intranuclear location of NF-ATc in endothelial cells is consistent with the previously demonstrated regulation of NF-

ATc activity in response to calcium fluxes in lymphocytes, where sustained increases in intracellular calcium levels result in activation of the calcium- and calmodulin-dependent phosphatase calcineurin. This protein is needed for the dephosphorylation and nuclear translocation of NF-AT⁶. Calcineurin is a target for the immunosuppressive drug cyclosporin A. To test whether this signalling pathway is present in cardiac endothelial cells, we collected wild-type embryos at e8.5, cultured them for 12 h in media, and then added cyclosporin A at 50 $\mu\text{g ml}^{-1}$. Embryos were cultured for an extra 5 or 20 h and three-channel confocal immunofluorescent microscopy was used to identify NF-ATc (monoclonal antibody 7A6, red fluorescence), PECAM-1 (monoclonal antibody 390, green fluorescence) and a nuclear stain (blue fluorescence). These studies (Fig. 4d, e) showed that embryos treated with cyclosporin A do not show nuclear localization of NF-ATc in cardiac endothelial cells. In untreated cultured embryos (Fig. 3d) and wild-type embryos *in vivo* (Fig. 4a, b), NF-ATc protein is primarily nuclear (pink staining), whereas in embryos treated with cyclosporin A most NF-ATc protein is cytoplasmic (Fig. 4e). As in NF-ATc mutants, inhibition of nuclear translocation of NF-ATc by cyclosporin A did not prevent initial endothelial cell activation, transformation to a mesenchymal phenotype (as shown by loss of PECAM-1 expression and migration into the extracellular matrix (Fig. 4d, e).

The heart and cardiovascular system are particularly sensitive to gene perturbations, with defects often leading to embryonic death^{7,11–13}. Cardiovascular system defects, most of which involve abnormal development of the cardiac valves and membranous septum¹⁴, occur in nearly 1% of newborns. However, null mutations in mice most often affect the formation of the compact ventricular myocardium^{15,16} and less frequently affect the trabecular myocardium^{17–19} and epicardium of the heart^{15,20}. The phenotype seen here is similar to that seen in *Sox-4*-deficient mice²¹. However, the spectrum of defects in the *Sox-4*^{-/-} mice suggests an abnormality in cranial neural crest function during early heart morphogenesis,

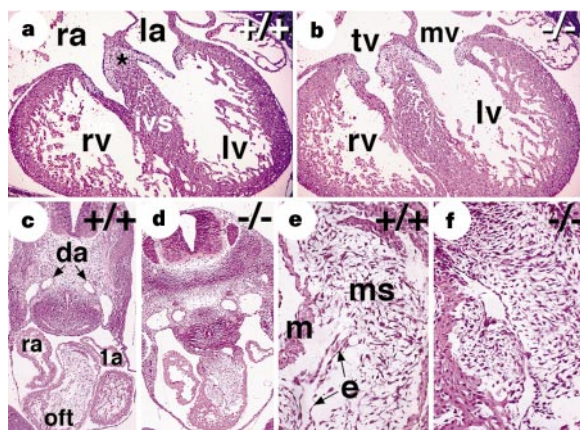


Figure 3 Histological analysis of e14.5 and e11.5 wild-type and NF-ATc^{-/-} embryos from line 141. Comparable cross-sections through the atrioventricular canal region of wild-type (a, c, e) and mutant (b, d, f) embryos are shown. Mutant embryos (b) have normal tricuspid and mitral valves. c–f, Evaluation of the endocardial cushions or conotruncal ridges of the outflow tract (oft) of e11.5 wild-type (c, e) and null-mutant (d, f) embryos. Ra, right atrium; rv, right ventricle; la, left atrium; lv, left ventricle; ivs, interventricular septum; mv, mitral valve; tv, tricuspid valve; da, dorsal aorta; nt, neural tube; ms, mesenchyme; m, myocardium; e, endothelial or endocardial lining of the outflow tract lumen.

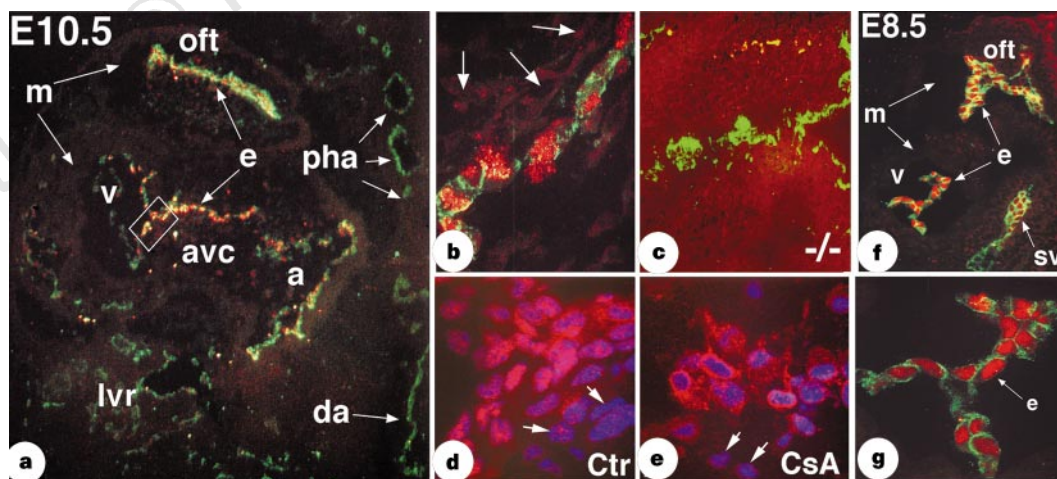


Figure 4 Developmental, temporal, and spatial expression of NF-ATc. a, Low-power, double-immunofluorescent, confocal photomicrograph of a sagittal section through the developing heart of an e10.5 wild-type embryo shows accentuation of NF-ATc expression (red, Cy5-labelled secondary antibody) within the outflow tract (oft) and atrioventricular canal region (avc) of the heart. Colocalization of NF-ATc expression with PECAM-1 (green, FITC-labelled secondary antibody) delineated a restriction of NF-ATc expression to the endothelium or endocardium (e) of the oft and avc. Attenuated expression was also seen within the endocardium of the ventricle (v) and atrium (a). Lvr, endothelium of the liver; da, dorsal aorta; pha, pharyngeal arch arteries; m, myocardium. b, A high-power magnification of the endocardium from the avc. NF-ATc expression was not detected in the endothelial cells that had undergone

mesenchymal transformation and had migrated into the endocardial cushions (indicated by arrows). c, Similar photomicrographs through the oft of a null-mutant e11.5 embryo. d, e, Cyclosporin A (CsA) blocks nuclear translocation of NF-ATc *in situ*. High-resolution, immunofluorescent confocal photomicrographs of cross-sections through the outflow tract of e10.0 control embryos (d) and embryos cultured in the presence of 50 $\mu\text{g ml}^{-1}$ of CsA for 20 h (e) are shown. NF-ATc expression is detected by a Cy5-conjugated secondary antibody (red). All cell nuclei are labelled with DAPI (blue). f, An e8.5 embryo showing NF-ATc expression restricted to the endocardium of the outflow tract, ventricle, and sinus venosus (sv; future right atrium). g, A high-power magnification of the ventricular endocardium reveals the localization of NF-ATc expression in the nucleus of the endothelial cells.

an abnormality which is unlikely in NF-ATc^{-/-} mice because of their circumscribed defect in semilunar valve morphogenesis. Nevertheless, the regional and temporal distribution of Sox-4 and NF-ATc and their obvious importance in outflow tract ontogeny justifies further investigation of potential interactions between these proteins.

Extensive semilunar valve insufficiency can cause *in utero* death in human embryos²². Thymic hypoplasia is also seen²², supporting the expression of NF-ATc in the embryonic thymus. These results may explain the significant increase in fetal death seen when cyclosporin A is administered to pregnant rats between e8 and e14 (ref. 23). Indeed, normal embryos cultured with cyclosporin A develop congestive heart failure (our unpublished observations). Examination of non-viable human embryos for NF-ATc gene mutations and caution in the use of cyclosporin A during pregnancy seem warranted.

Similar mechanisms are thought to regulate the formation of the semilunar and atrioventricular valves⁸. However, the selective defects in the semilunar valves seen here imply the existence of some essential mechanisms that are unique to aortic and pulmonary valve formation. There is differential gene expression in the outflow and inflow portions of the heart²⁴ and use of an insertion mutation has identified a locus on mouse chromosome 13 that is required for normal conotruncal development²⁵. However, all of these genes are primarily expressed by the myocardium. Our results are the first to indicate distinct, endothelial-dependent, molecular pathways controlling semilunar valve formation. This conclusion concurs with the clinical spectrum of congenital heart disease: where it is rare to detect simultaneous defects in atrioventricular and semilunar valve formation²⁶.

The restriction of NF-ATc expression to the endothelial lining of the heart is unique in development, as NF-ATc is the only endothelial-specific, and only the second endothelial-specific²⁷, transcription factor to be described. Selective endothelial proliferation on the arterial face of the valve cusps²⁸, leading to semilunar valve formation, may result from an 'inductive' interaction between the endocardium and the underlying mesenchyme of the endocardial cushions. In an *in vitro* model system designed to observe the invasion of atrioventricular canal endothelium into a collagen gel^{8,29}, endothelial 'transformation' was shown to depend upon increases in levels of intracellular free calcium³⁰. One possible explanation for the phenotype seen might be that NF-ATc mediates calcium-dependent initiation of transformation. However, this is not the case as NF-ATc^{-/-} endothelial cells delaminated from the endocardial layer and populated the endocardial-cushion extracellular matrix (Figs 3, 4). Thus, NF-ATc must be involved in events that are essential for valve formation after, and independently of endothelial transformation. These events do not involve the production of transforming growth factor β proteins 1–3, which are expressed at normal levels in NF-ATc^{-/-} heart (our unpublished observations). Future experiments will focus on identifying the upstream targets and downstream effectors of NF-ATc. □

Methods

Disruption of the NF-ATc gene. An NF-ATc genomic clone was isolated by screening a 129/sv genomic library with a complementary DNA fragment corresponding to the NF-ATc Rel-homology domain. A 2-kilobase (kb) fragment from this clone was replaced with a thymidine kinase (TK) promoter neo-poly(A) cassette from pMC1neopoly(A) (Stratagene) by transferring a PstI–NdeI fragment containing 6 kb of 5' NF-ATc sequence, the neo cassette, and 2 kb of 3' NF-ATc sequence to a plasmid encoding the TK drug-resistance gene. ES cell clones transfected with the targeting construct were selected for drug resistance in 180 μ g ml⁻¹ G418 and 2 μ M gancyclovir. Homologous recombinants were identified by digesting genomic DNA from drug-resistant clones with BamHI and hybridizing with the 0.8 kb Asp 718/BamHI 3' probe. Offspring were genotyped by Southern blot analysis of tail or yolk-sac DNA. Two different primers, indicated in Fig. 1, were used to amplify a 275-base-pair

(bp) fragment from the wild-type allele or a 350-bp fragment from the line 141 mutant allele.

Histological analysis and immunohistochemistry. Whole embryos were fixed in buffered formalin or 4% paraformaldehyde and embedded in paraffin. Sagittal and transverse sections were cut, and stained with hematoxylin–eosin. For immunohistochemistry, embryos were snap-frozen in optimal cutting temperature (OCT) compound (Miles Laboratories); sections (of 5–6 μ m) were thaw-mounted on 0.05% poly-L-lysine-coated slides and post-fixed with 1% paraformaldehyde. Tissue sections were pretreated with 2.5% goat serum and incubated overnight at 4 °C with purified tissue culture supernatant of a rat monoclonal antibody against murine PECAM-1 at 25 μ g ml⁻¹ in Tris-buffered saline (TBS) and a monoclonal antibody ascites (7A6; a gift from G. Crabtree) specific for the amino terminus of NF-ATc at 1:250 dilution. A control isotype antibody and no primary antibody were used as controls. Sections were then rinsed in TBS and incubated for 1 h in a mixture of fluorescein isothiocyanate (FITC)-conjugated goat secondary antibody against a rat protein (1:200 dilution) preabsorbed against mouse antigens, and a biotin-conjugated goat secondary antibody against mouse IgG1 (1:200 dilution, Southern Biotechnical). Sections were rinsed in TBS, incubated with Cy5-conjugated streptavidin (Jackson ImmunoResearch) for 15 min, washed and mounted in vectashield (Vector) anti-quench media with DAPI at a concentration of 0.75 μ g ml⁻¹.

Confocal microscopy. Immunostained specimens were optically sectioned using a computer-interfaced, laser scanning microscope (Leica TCS 4D). This equipment is fitted with a 488 nm/568 nm/647 nm krypton–argon laser, allowing, simultaneous analysis of fluorescein and rhodamine chromophores, with the option of rescanning for Cy5 fluorescence (in the presence of a 600-nm band-pass filter) and ultraviolet-excitable fluorochromes. Images were acquired at a resolution of 1,024 $\times$ 1,024 pixels and compiled on an image processing (Os9) workstation; subsequent figures were composed on a Macintosh 7600.120 PowerPC using Photoshop 4.0 (Adobe).

Whole-embryo culture. Postimplantation mouse embryos were cultured as described³¹. Briefly, embryos collected from ICR mice at e8.5 were dissected free from the decidua and Reichardt's membrane; the yolk sac was left intact and cultured (three embryos per 10 ml media) in 60% serum and 40% Tyrodes at an initial O₂ concentration of 20% for 12 h, followed by increasing concentrations of O₂ to 40% at 24 h and 90% after 36 h with 5% CO₂ and balance N₂. Cyclosporin A was added at 50 μ g ml⁻¹.

Received 16 September; accepted 22 December 1997.

- McCaffrey, P. G. *et al.* Isolation of the cyclosporin-sensitive T cell transcription factor NFATp. *Science* **262**, 750–754 (1993).
- Rao, A., Luo, C. & Hogan, P. G. Transcription factors of the NFAT family: regulation and function. *Annu. Rev. Immunol.* **15**, 707–747 (1997).
- Hoey, T., Sun, Y.-L., Williamson, K. & Xu, X. Isolation of two new members of the NF-AT gene family and functional characterization of the NF-AT proteins. *Immunity* **2**, 461–472 (1995).
- Northrop, J. P. *et al.* NF-AT components define a family of transcription factors targeted in T-cell activation. *Nature* **369**, 497–502 (1994).
- Emmel, E. A. *et al.* Cyclosporin A specifically inhibits function of nuclear proteins involved in T cell activation. *Science* **246**, 1617–1620 (1989).
- Timmerman, L. A., Clipstone, N. A., Ho, S. N., Northrop, J. P. & Crabtree, G. R. Rapid shuttling of NF-AT in discrimination of Ca²⁺ signals and immunosuppression. *Nature* **383**, 837–840 (1996).
- Rossant, J. Mouse mutants and cardiac development. *Circ. Res.* **78**, 349–353 (1996).
- Eisenburg, L. M. & Markwald, R. R. Molecular regulation of atrioventricular valvuloseptal morphogenesis. *Circ. Res.* **77**, 106 (1995).
- Kwee, L. *et al.* Defective development of the embryonic and extraembryonic circulatory systems in vascular cell adhesion molecule (VCAM-1) deficient mice. *Development* **121**, 489–503 (1995).
- Baldwin, H. S., Jensen, K. L. & Solursh, M. Myogenic cytodifferentiation of the precardiac mesoderm in the rat. *Differentiation* **47**, 163–172 (1991).
- Copp, A. J. Death before birth: clues from gene knockouts and mutations. *Trends Genet.* **11**, 87–93 (1995).
- Olson, E. N. & Srivastava, D. Molecular pathways controlling heart development. *Science* **272**, 671–676 (1996).
- Fishman, M. C. & Chien, K. R. Fashioning the vertebrate heart: earliest embryonic decisions. *Development* **124**, 2099–2177 (1997).
- Ferencz, C. *et al.* Congenital heart disease: prevalence at live birth. *Am. J. Epidemiol.* **121**, 31–36 (1985).
- Kwee, L. *et al.* Defective development of the embryonic and extraembryonic circulatory systems in vascular cell adhesion molecule (VCAM-1) deficient mice. *Development* **121**, 489–503 (1995).
- Moens, C. B., Stanton, B. R., Parada, L. F. & Rossant, J. Defects in heart and lung development in compound heterozygotes for two different targeted mutations of the N-myc locus. *Development* **119**, 485–499 (1993).
- Meyer, D. & Birchmeier, C. Multiple essential functions of neuregulin in development. *Nature* **23**, 386–390 (1995).
- Gassmann, M. *et al.* Aberrant neural and cardiac development in mice lacking the ErbB4 neuregulin receptor. *Nature* **378**, 390–394 (1995).
- Lee, K. F. *et al.* Requirement for neuregulin receptor erbB2 in neural and cardiac development. *Nature* **378**, 394–398 (1995).
- Yang, J. T., Rayburn, H. & Hynes, R. O. Cell adhesion events mediated by α 4 integrins are essential in placental and cardiac development. *Development* **121**, 549–560 (1995).

21. Schilham, M. W. *et al.* Defects in cardiac outflow tract formation and pro-B-lymphocyte expansion in mice lacking Sox-4. *Nature* **380**, 711–714 (1996).
22. Miyabara, S. *et al.* Absent aortic and pulmonary valves: investigation of three fetal cases with cystic hygroma and review of the literature. *Heart Vessels* **9**, 49–55 (1994).
23. Brown, P. A. J., Gray, E. S., Whiting, P. H., Simpson, J. G. & Thomson, A. W. Effects of cyclosporin A on fetal development in the rat. *Biol. Neonate* **48**, 172–180 (1985).
24. Lyons, K. M., Hogan, B. L. & Robertson, E. J. Colocalization of BMP-7 and BMP-2 RNAs suggest that these factors cooperatively mediate tissue interactions during murine development. *Mech. Dev.* **50**, 71–83 (1995).
25. Yamamura, H., Zhang, M., Markwald, R. R. & Mjaatvedt, C. H. A heart segmental defect in the anterior-posterior axis of a transgenic mutant mouse. *Dev. Biol.* **186**, 58–72 (1997).
26. Clark, E. B. In *Genetics of Cardiovascular Disease* (eds Pierpont, M. E. & Moller, J. H.) 3–11 (Martinus-Nijhoff, Boston, Massachusetts, 1987).
27. Tian, H., McKnight, S. L. & Russell, D. W. Endothelial PAS domain protein 1 (EPAS1), a transcription factor selectively expressed in endothelial cells. *Genes Dev.* **11**, 72–82 (1997).
28. Hurler, J. M., Colvée, E. & Blanco, A. M. Development of mouse semilunar valves. *Anat. Embryol.* **160**, 83–91 (1980).
29. Markwald, R. R. *et al.* In *The Role of Extracellular Matrix in Development* (ed. Trelstead, R. L.) 323–350 (Allan R. Liss Inc, New York, 1994).
30. Runyan, R. B. *et al.* Signal transduction of a tissue interaction during embryonic heart development. *Cell Regulation* **1**, 301–313 (1990).
31. Sturm, K. & Tam, P. P. L. *Methods Enzymol.* **225**, 164–190 (1993).

Acknowledgements. We thank G. Crabtree and L. Timmerman for NF-AT reagents; E. Robertson for help in embryo dissection and preparation; C. Freedman for preparation of the manuscript; and P. Bannerman and T. Oliver of the Children's Hospital of Philadelphia and University of Pennsylvania Cancer Center Confocal Microscopy Core for advice and assistance. This work was supported by grants from the NIH (L.H.G. and H.S.B.), a National Science Foundation Fellowship (A.M.R.) and a gift from the G. Harold and Leila Y. Mathers Charitable Foundation (L.H.G.). M.J.G. is a scholar of the Leukemia Society of America. H.S.B. is an Established Investigator of the American Heart Association.

Correspondence and requests for materials should be addressed to L.H.G. (e-mail: lglmiche@hsph.harvard.edu).

A causal role for E-cadherin in the transition from adenoma to carcinoma

Anne-Karina Perl¹*, Petra Wilgenbus¹*, Ulf Dahl¹†, Henrik Semb²† & Gerhard Christofori¹*

¹ Research Institute of Molecular Pathology, Dr Bohr-Gasse 7, A-1030 Vienna, Austria

² Department of Microbiology, Umeå University, S-90187 Umeå, Sweden

Development of malignant tumours is in part characterized by the ability of a tumour cell to overcome cell–cell adhesion and to invade surrounding tissue. E-cadherin is the main adhesion molecule of epithelia^{1–3}, and it has been implicated in carcinogenesis because it is frequently lost in human epithelial cancers^{4–6}. Re-establishing the functional cadherin complex in tumour cell lines results in a reversion from an invasive to a benign epithelial phenotype⁷. However, it remained unresolved whether the loss of E-cadherin-mediated cell adhesion was a cause or a consequence of tumour progression *in vivo*. Here we report that the loss of E-cadherin expression coincides with the transition from well differentiated adenoma to invasive carcinoma in a transgenic mouse model of pancreatic β -cell carcinogenesis (Rip1Tag2)⁸. Intercrossing Rip1Tag2 mice with transgenic mice that maintain E-cadherin expression in β -tumour cells results in arrest of tumour development at the adenoma stage, whereas expression of a dominant-negative form of E-cadherin induces early invasion and metastasis. The results demonstrate that loss of E-cadherin-mediated cell adhesion is one rate-limiting step in the progression from adenoma to carcinoma.

We have studied tumour progression in a transgenic mouse line that expresses the SV40 T antigen under the control of the insulin promoter in the β -cells of pancreatic islets of Langerhans (Rip1Tag2)⁸. These mice reproducibly develop tumours in a multistage tumorigenesis pathway involving islet hyperplasia (~70% of all islets)^{9,10}, the formation of new blood vessels (angiogenesis, ~20%)^{11,12} and, finally, solid β -cell tumours (insulinomas) that can be further distinguished into well differentiated,

benign tumours (adenomas, ~2%) and de-differentiated, invasive tumours (carcinomas, ~0.5%). In β -cell adenomas, tumour cells are arranged as epithelial structures in cords and ribbons. The tumour's encapsulation, the lack of local invasion, as well as the relatively normal nuclear morphology and nuclear to cytoplasmic ratio define these tumours as benign (Fig. 1A, a and b). In contrast, β -cell carcinomas are recognized by their local invasion, by remnants of successive fibrous capsules that by expansive growth are embedded in the mass, by cellular anaplasia, nuclear atypia, and high nuclear to cytoplasmic ratio (Fig. 1Ac). Metastases are usually not found in these mice, probably because they succumb to hypoglycaemia with increased tumour mass.

Immunohistochemical analysis of the different stages of Rip1Tag2 tumorigenesis reveals that E-cadherin is expressed at the surface of β -cells in normal islets of Langerhans and in the acini cells of the exocrine pancreas (Fig. 1Ad). E-cadherin expression is unaltered in adenoma cells (Fig. 1Ae). In contrast, E-cadherin is completely absent in all carcinomas examined (Fig. 1Af; >200 tumours were analysed). Notably, in a subset of tumours E-cadherin expression is focally lost concomitant with the loss of benign hallmarks (data not shown), indicating that the expression of E-cadherin is downregulated with the transition from β -cell adenoma to β -cell carcinoma. This loss of E-cadherin expression is due to transcriptional repression (ref. 13; data not shown).

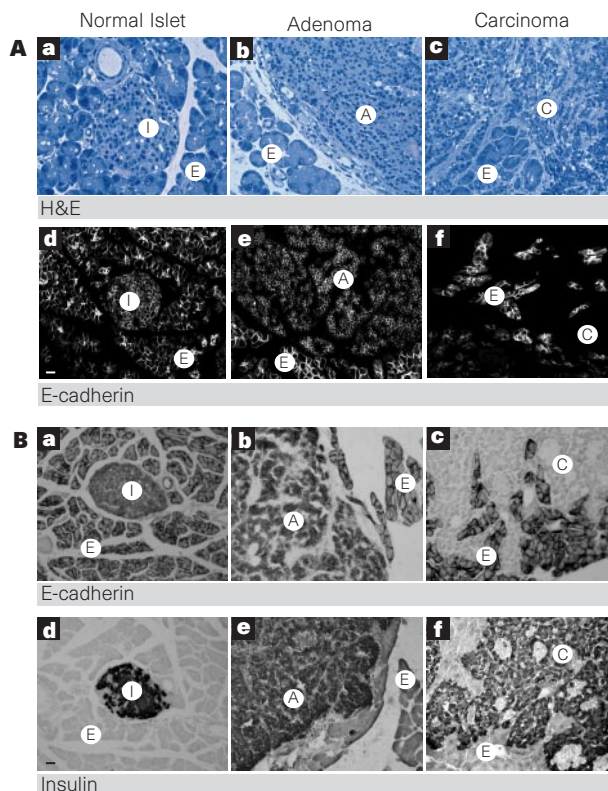


Figure 1 Histopathological and immunohistochemical analyses of tumorigenesis. **A**, Transition from β -cell adenoma to β -cell carcinoma in Rip1Tag2 transgenic mice. **a–c**, Histopathological analysis of β -cell tumour development by haematoxylin/eosin staining (H&E). **d–f**, Immunofluorescence analysis of E-cadherin expression in the different stages of β -cell tumorigenesis. **B**, E-cadherin-negative carcinoma cells are insulin-producing β -cells. **a–c**, Immunohistochemical analysis of E-cadherin expression in the multiple stages of β -cell tumorigenesis. **d–f**, Immunohistochemical analysis of insulin expression on adjacent sections to **a–c**. A, Adenoma; C, carcinoma; E, exocrine pancreas; I, normal islet of Langerhans. Scale bars, 25 μ m.

The tumour cells that have lost E-cadherin expression at the carcinoma stage still express insulin, indicating that they are indeed β -tumour cells (Fig. 1B). However, insulin expression is absent in a subset of highly anaplastic carcinomas (~30% of the carcinomas) suggesting a process of de-differentiation. Expression of two more members of the cadherin family, N-cadherin (Fig. 2) and R-cadherin (data not shown) is not downregulated with the transition from adenoma to carcinoma. Furthermore, the epithelial polarity marker Na/K-ATPase (Fig. 2) and the tight junction protein ZO1 (data not shown) are only modestly downregulated or disorganized in the invasive tumours. No change in the organization of F-actin and fibronectin is apparent, and vimentin, a marker for mesenchymal cells, is not expressed in the carcinoma stages (data not shown), indicating that the transition from adenoma to carcinoma in Rip1Tag2 transgenic mice does not represent a complete epithelial-mesenchymal conversion. Of the proteins analysed, loss of E-cadherin expression is the only notable change in this process.

To assess whether the loss of E-cadherin-mediated cell adhesion is required for tumour progression *in vivo* we sought to maintain expression of E-cadherin throughout β -cell tumorigenesis in

Rip1Tag2 transgenic mice. Five independent transgenic mouse lines were generated that specifically expressed E-cadherin under the control of the rat insulin promoter in pancreatic β -cells (Rip1E-cad). None of them exhibited abnormalities in islet development and islet physiology (data not shown). All five Rip1E-cad mouse lines were intercrossed with Rip1Tag2 transgenic mice, and changes in tumour development in single-transgenic (Rip1Tag2) and double-transgenic (Rip1Tag2 $\times$ Rip1E-cad) mice within the same litters were monitored. In single-transgenic mice, β -cell tumours developed without any significant aberrations from the well studied pathway (see above). In double-transgenic mice, tumour volumes were moderately reduced, whereas the number of tumours per mouse did not significantly change (Table 1). Histopathological analysis revealed that the incidence of carcinoma formation was reduced from 26.5% of all tumours in single-transgenic mice to 7.8% in double-transgenic mice (Table 1; $P < 0.001$, chi-square test). Notably, the few carcinomas in the double-transgenic mice had all lost E-cadherin protein expression, including expression from the transgene (Fig. 3), indicating that β -tumour cells have developed a strong selection against E-cadherin-mediated cell adhesion. The downregulation of E-cadherin transgene expression might be due to destabilization of E-cadherin message or protein, because neither the transgene was lost nor insulin promoter activity was downregulated in these tumour cells (data not shown). These

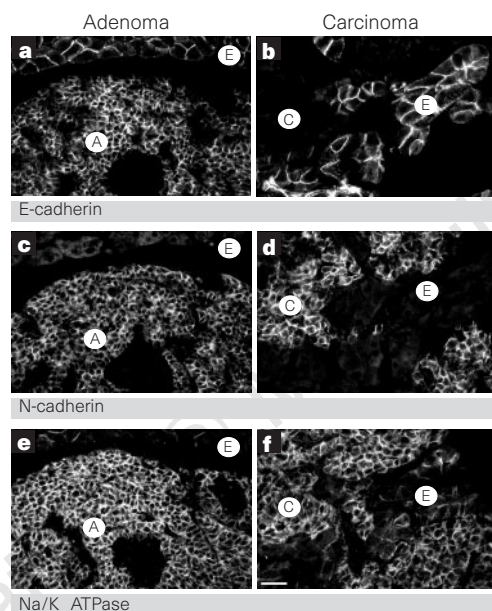


Figure 2 Expression of epithelial markers during the transition from β -cell adenoma to carcinoma in Rip1Tag2 transgenic mice. Immunofluorescent staining of adjacent histological sections from adenoma (a, c, e) or carcinoma (b, d, f) with antibodies against E-cadherin (a, b), N-cadherin (c, d) or Na/K-ATPase (e, f). A, Adenoma; C, carcinoma; E, exocrine pancreas. Scale bar, 25 μ m.

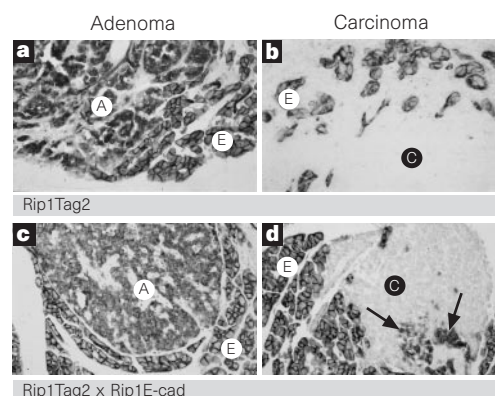


Figure 3 Forced expression of E-cadherin blocks the transition from adenoma to carcinoma. Immunohistochemical analysis of E-cadherin expression in adenomas (a, c) and carcinomas (b, d) of single-transgenic Rip1Tag2 (a, b) and double-transgenic Rip1Tag2 $\times$ Rip1E-cad mice (c, d). The few carcinomas found in Rip1Tag2 $\times$ Rip1E-cad double-transgenic mice (d) have focally lost E-cadherin expression, including expression from the E-cadherin transgene (β -tumour cells that still express E-cadherin are indicated by arrows). A, Adenoma; C, carcinoma; E, exocrine pancreas. Magnification: $\times 200$.

Table 1 Impact of E-cadherin function on β -cell tumorigenesis *in vivo*

Genotype	Tumour incidence* (per mouse)	Tumour volume* (mm ³)	Tumour phenotype†			
			Number of tumours analysed	Adenoma	Carcinoma	Metastasis
Rip1Tag2‡	7.0 $\pm$ 4.9	<12w: 32.3 $\pm$ 27.4 >12w: 58.7 $\pm$ 39.3	121	73.5%	26.5%	–
Rip1Tag $\times$ Rip1dnE-cad‡	6.0 $\pm$ 3.0	<12w: 20.2 $\pm$ 20.1 >12w: 26.4 $\pm$ 22.6	103	92.2%	7.8%	–
Rip1Tag2§	5.5 $\pm$ 3.6	12w: 42.2 $\pm$ 31.8	49	75.5%	24.5%	–
Rip1Tag $\times$ Rip1dnE-cad§	7.5 $\pm$ 2.6	12w: 48.9 $\pm$ 14.1¶	85	45.9%	50.6%	3.5%

* Tumours of a diameter of more than 1 mm were scored. Total tumour volumes were calculated as a function of an approximate spherical shape.

† Based on histopathological analysis.

‡ Mice were analysed between 10 and 18 weeks of age and grouped into mice younger than 12 weeks (<12w) and mice older than 12 weeks (>12w) as indicated.

§ Mice were analysed at the age of 12 weeks.

|| Note that the carcinomas found in Rip1Tag2 $\times$ Rip1E-cad double-transgenic mice have completely lost E-cadherin expression, including expression from the E-cadherin transgene.

¶ Only primary tumours were included.

results indicate that loss of E-cadherin expression is a rate-limiting step in the progression to the carcinoma stage *in vivo*.

Next we investigated whether loss of E-cadherin function is sufficient to promote tumour invasion *in vivo*. Rip1Tag2 mice were intercrossed with transgenic mice that carry a dominant-negative form of E-cadherin, lacking the extracellular domain, under the control of the insulin promoter (Rip1dnE-cad)¹⁴. Expression of the truncated form of E-cadherin in normal mice displaces both E- and N-cadherin from pancreatic β -cells resulting in a perturbation of β -cell aggregation during islet development¹⁴. However, in adult mice normal islets form, most likely because N-cadherin, in contrast to E-cadherin, is no longer displaced from β -cells after birth (U.D. and H.S., unpublished results). Tumour incidence or tumour volume were not significantly changed between double-transgenic Rip1Tag2 $\times$ Rip1dnE-cad mice and single-transgenic Rip1Tag2 littermates. However, the incidence of carcinoma formation was dramatically increased from 24.5% of all tumours in single-transgenic mice to 50.6% in double-transgenic mice (Table 1; $P < 0.001$, chi-square test). In addition, one fourth of the double-transgenic mice developed metastases to the pancreatic lymph nodes (3.5% of solid tumours), an invasive phenotype that was never observed in single-transgenic Rip1Tag2 mice^{8–12} (Table 1). Histopathological analysis reveals two types of invasive tumours in the Rip1Tag2 $\times$ Rip1dnE-cad double-transgenic mice (Fig. 4): tumours that clearly overgrew their capsule and invaded into the surrounding tissue, yet exhibit a relatively normal cellular morphology (referred to as early carcinoma), and invasive tumours that exhibit all hallmarks of highly malignant tumour cells, including nuclear atypia and cellular anaplasia (referred to as late carcinoma). Some of these invasive tumours contain tumour cells of both normal and malignant cellular phenotype indicating a gradual

early-late transition during tumour progression (Fig. 4e). The lymph node metastases of double-transgenic mice were all of the late carcinoma phenotype (Fig. 4). Insulin expression indicates that the carcinomas and the lymph node metastases have originated from β -cell tumours (Fig. 4). In contrast to the carcinomas of single-transgenic Rip1Tag2 mice, endogenous E-cadherin is still expressed in the invasive carcinoma stages of double-transgenic Rip1Tag2 $\times$ Rip1dnE-cad mice (Fig. 4). The dominant-negative form of E-cadherin appears to induce subcellular redistribution and destabilization of endogenous E-cadherin in late carcinoma and in metastasis (Fig. 4). Taken together, abrogation of cadherin-mediated cell adhesion is sufficient to induce early tumour invasion and metastasis.

The present findings provide the first direct evidence for a causal role of E-cadherin in the progression from benign adenoma to malignant carcinoma *in vivo*. The data suggest that E-cadherin counteracts the tumour cell's transition to an invasive phenotype by maintaining intercellular adhesion and epithelial organization. E-cadherin-mediated cell adhesion has no direct effect on the control of tumour cell proliferation; the incidence of both mitosis and apoptosis does not differ significantly between the different tumour phenotypes (mitosis: $22 \pm 5.6\%$ for adenoma versus $18.3 \pm 5.3\%$ for carcinoma; apoptosis: $0.43 \pm 0.18\%$ for adenoma versus $0.32 \pm 0.12\%$ for carcinoma). These results are surprising because the mere loss of cell adhesion is not expected to enable cells to cross the basal lamina and to invade surrounding tissue. It is likely that the loss of E-cadherin-mediated cell adhesion may convey additional signals to induce tumour-cell invasion. Recently, it has been demonstrated that β -catenin, in addition to its role in cadherin-mediated cell adhesion, is part of the Wnt signalling pathway^{15–17}. Cytosolic β -catenin is translocated to the nucleus, where it binds Tcf/Lef-1 transcription factors and associates with a number of promoter sequences and potentially affects transcription^{18–20}. Activation of this pathway by mutation of either β -catenin or APC has been shown to play an important role in the development of colon carcinoma and melanoma^{21–23}. The observation that β -catenin can transduce signals to the nucleus raises the possibility that changes in E-cadherin-mediated cell adhesion may modulate gene expression and thus tumour phenotype. Furthermore, E-cadherin's transcriptional repression in human cancers and in Rip1Tag2 transgenic mice needs to be investigated. Chromatin rearrangement, hypermethylation, and loss of transcription factor binding coincide with suppression of E-cadherin promoter activity in invasive carcinoma cells^{24–26}; however, the molecular signals responsible remain unknown. Thus, unravelling the regulation of E-cadherin expression and the communication pathway between cell adhesion and gene expression may lead to a better understanding of how tumour cells control their invasive phenotype. □

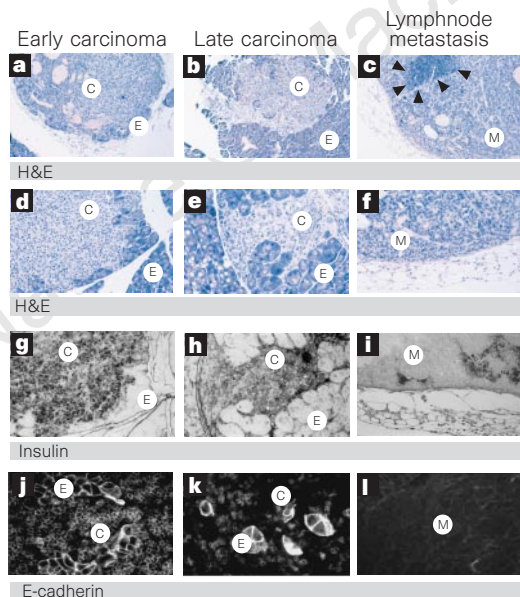


Figure 4 Forced expression of dominant-negative E-cadherin induces tumour-cell invasion and metastasis. **a–c**, H&E staining of histological sections containing early carcinoma, late carcinoma, and lymph node metastasis from Rip1Tag2 $\times$ Rip1dnE-cad mice as indicated. Arrowheads mark lymphoid tissue that has been overgrown by metastatic β -tumour cells. **d–f**, Higher magnification microphotographs of H&E staining of different tumour stages in double-transgenic Rip1Tag2 $\times$ Rip1dnE-cad mice as indicated. **g–i**, Immunohistochemical analysis of insulin expression on sections adjacent to **d–f**. **j–l**, Immunohistochemical analysis of endogenous E-cadherin expression in the different tumour stages using antibodies specific for the extracellular domain of E-cadherin. C, Carcinoma; E, exocrine pancreas; M, metastasis. Magnifications: **a–c**, $\times 100$; **d–l**, $\times 200$.

Methods

Transgenic mice. The generation and phenotypic characterization of Rip1Tag2^{8–12} and Rip1dnE-cad¹⁴ transgenic mice has been described previously. Rip1E-cad transgenic mice were generated according to standard procedures²⁷. The transgene was generated by inserting a 3.0 kilobase (kb) *XbaI/HindIII* complementary DNA fragment encoding mouse E-cadherin (a gift from M. Takeichi, Kyoto) into the plasmid pRip1Tag⁸ from which the fragment encoding the SV40 early region was removed by digestion with *XbaI* and *HindIII*. Five independent founder lines were found to express the transgene in pancreatic β -cells and were subsequently backcrossed to C57/BL6 for at least 2 generations.

Tumour volumes. Glucose (5%, w/v) was added to the drinking water of all tumour-bearing mice to counteract hypoglycaemia resulting from insulinoma development. Mice were killed between 10 and 16 weeks of age, tumour numbers and diameters were determined, and the tumour volume was calculated assuming a spherical tumour.

DNA synthesis and apoptosis assays. For BrdU staining, tissues were prepared as above, except that mice were injected intraperitoneally with 100 μ g of bromodeoxyuridine (BrdU) per gram of body weight 2 hours before being

killed. Histological sections were immunostained as previously described²⁸. The mitotic index equals the percentage of BrdU-positive nuclei. Detection of apoptotic cells using the TUNEL-technique was performed as previously described²⁸. The apoptotic index is the percentage of TUNEL-positive nuclei.

Histopathological analysis. Tissue embedding in JB4 plastic and preparation of semi-thin sections was as described¹⁰. For immunohistochemical analysis pancreas were removed and fixed in HBS-Ca²⁺ (HEPES-buffered saline, 1 mM CaCl₂) containing 4% paraformaldehyde for 2 h at 4°C¹⁴. Tissues were incubated overnight at 4°C in HBS-Ca²⁺/30% sucrose, embedded in O.C.T. compound (Tissue Tek), and frozen in liquid nitrogen. Immunostaining was as described¹⁴. Specific binding of primary antibodies was visualized using either the ABC Vector-horseradish peroxidase kit according to the manufacturer's recommendations or fluorescence-labelled secondary antibodies as indicated. Antibodies were monoclonal rat anti-mouse E-cadherin (Zymed; 1:150), monoclonal rat anti-mouse N-cadherin¹⁴ (a gift from M. Takeichi, Kyoto), polyclonal guinea pig anti-insulin (DAKO, 1:150), and rabbit polyclonal serum against mouse Na/K-ATPase (a gift from W. J. Nelson, Stanford, 1:500). Biotin-conjugated donkey anti-rat and anti-rabbit IgG (1:500) and Cy3-streptavidin (1:750) were purchased from Jackson ImmunoResearch Laboratories.

Received 18 September; accepted 21 November 1997.

- Geiger, B. & Ayalon, O. Cadherins. *Annu. Rev. Cell Biol.* **8**, 307–332 (1992).
- Takeichi, M. Morphogenetic roles of classic cadherins. *Curr. Opin. Cell Biol.* **7**, 619–627 (1995).
- Aberle, H., Schwartz, H. & Kemler, R. Cadherin-catenin complex: protein interactions and their implications for cadherin function. *J. Cell. Biochem.* **61**, 514–523 (1996).
- Birchmeier, W., Weidner, K. M., Hülsken, J. & Behrens, J. Molecular mechanisms leading to cell junction (cadherin) deficiency in invasive carcinomas. *Sem. Cancer Biol.* **4**, 231–218 (1993).
- Takeichi, M. Cadherins in cancer: implications for invasion and metastasis. *Curr. Opin. Cell Biol.* **5**, 806–811 (1993).
- Birchmeier, W. & Behrens, J. Cadherin expression in carcinomas: role in the formation of cell junctions and the prevention of invasiveness. *Biochim. Biophys. Acta* **1198**, 11–26 (1994).
- Vlemminckx, K., Vakaet, L., Mareel, M., Fiers, W. & Van Roy, F. Genetic manipulation of E-cadherin expression by epithelial tumor cells reveals an invasion suppressor role. *Cell* **66**, 107–119 (1991).
- Hanahan, D. Heritable formation of pancreatic beta-cell tumours in transgenic mice expressing recombinant insulin/simian 40 oncogenes. *Nature* **315**, 115–122 (1985).
- Teitelman, G., Alpert, S. & Hanahan, D. Proliferation, senescence, and neoplastic progression of β cells in hyperplastic pancreatic islets. *Cell* **52**, 97–105 (1988).
- Christofori, G., Naik, P. & Hanahan, D. A second signal supplied by insulin-like growth factor II in oncogene-induced tumorigenesis. *Nature* **369**, 414–418 (1994).
- Folkman, J., Watson, K., Ingber, D. & Hanahan, D. Induction of angiogenesis during the transition from hyperplasia to neoplasia. *Nature* **339**, 58–61 (1989).
- Christofori, G., Naik, P. & Hanahan, D. Vascular endothelial growth factor and its receptors, flt-1 and flk-1, are expressed in normal pancreatic islets and throughout islet cell tumorigenesis. *Mol. Endocrinol.* **9**, 1760–1770 (1995).
- Hutton, J. C. *et al.* Molecular cloning of mouse pancreatic islet R-cadherin: differential expression in endocrine and exocrine tissue. *Mol. Endocrinol.* **7**, 1151–1160 (1993).
- Dahl, U., Sjödin, A. & Semb, H. Cadherins regulate aggregation of pancreatic β -cells in vivo. *Development* **122**, 2895–2902 (1996).
- Huber, O., Bierkamp, C. & Kemler, R. Cadherins and catenins in development. *Curr. Opin. Cell Biol.* **8**, 685–691 (1996).
- Miller, J. R. & Moon, R. T. Signal transduction through β -catenin and specification of cell fate during embryogenesis. *Genes Dev.* **10**, 2527–2539 (1996).
- Gumbiner, B. M. Carcinogenesis: a balance between β -catenin and APC. *Curr. Biol.* **7**, 443–446 (1997).
- Molenaar, M. *et al.* XTcf-3 transcription factor mediates β -catenin-induced axis formation in *Xenopus* embryos. *Cell* **86**, 391–399 (1996).
- Behrens, J. *et al.* Functional interaction of β -catenin with the transcription factor LEF-1. *Nature* **382**, 638–642 (1996).
- Huber, O. *et al.* Nuclear localization of β -catenin by interaction with transcription factor LEF-1. *Mech. Dev.* **59**, 3–10 (1996).
- Korinek, V. *et al.* Constitutive transcriptional activation by a β -catenin-Tcf complex in APC^{-/-} colon carcinoma. *Science* **275**, 1784–1787 (1997).
- Morin, P. J. *et al.* Activation of β -catenin-Tcf signaling in colon cancer by mutations in β -catenin or APC. *Science* **275**, 1787–1790 (1997).
- Rubinfeld, B. *et al.* Stabilization of β -catenin by genetic defects in melanoma cell lines. *Science* **275**, 1790–1792 (1997).
- Hennig, G., Löwrick, O., Birchmeier, W. & Behrens, J. Mechanisms identified in the transcriptional control of epithelial gene expression. *J. Biol. Chem.* **271**, 595–602 (1996).
- Yoshiura, K. *et al.* Silencing of the E-cadherin invasion-suppressor gene by CpG methylation in human carcinomas. *Proc. Natl Acad. Sci. USA* **92**, 7416–7419 (1995).
- Graff, J. R. *et al.* E-cadherin expression is silenced by hypermethylation in human breast and prostate carcinomas. *Cancer Res.* **55**, 5195–5199 (1995).
- Hogan, B. L. M., Constantini, F. & Lacey, E. *Manipulating the Mouse Embryo: a Laboratory Manual* (Cold Spring Harbor Laboratory Press, New York, 1986).
- Naik, P., Karrim, J. & Hanahan, D. The rise and fall of apoptosis during multistage tumorigenesis: down-modulation contributes to tumor progression from angiogenic progenitors. *Genes Dev.* **10**, 2105–2116 (1996).

Acknowledgements. We thank D. Hanahan and R. Kelly for support and discussions. We are grateful to W. Jochum for expertise in histopathology and M. Herzog and S. Luef for technical assistance. We thank M. Cotten, E. F. Wagner, H. Beug, K. Nasmyth and G. M. Lamm for critical comments on the manuscript. Animal care was in accordance with institutional guidelines. Supported in part by the Austrian Industrial Research Promotion Fund (A.-K.P., P. W., G.C.) and by the Swedish Cancer Society, Lion's Cancer Research Foundation, Umeå University and M. Gergylls Stiftelse (U.D., H.S.).

Correspondence and requests for materials should be addressed to G.C. (e-mail: christofori@nt.imp.univie.ac.at).

A lipid associated with the antiphospholipid syndrome regulates endosome structure and function

Toshihide Kobayashi*, Espen Stang†, Karen S. Fang*, Philippe de Moerloose‡, Robert G. Parton† & Jean Gruenberg*

* Department of Biochemistry, Sciences II, 30 quai E. Ansermet, 1211-Geneva-4, Switzerland

† Centre for Microscopy and Microanalysis, Department of Physiology and Pharmacology, and Centre for Molecular and Cellular Biology, University of Queensland, Queensland 4072, Australia

‡ Division of Angiology and Haemostasis, University Hospital of Geneva, 1211-Geneva-14, Switzerland

Little is known about the structure and function of membrane domains in the vacuolar apparatus of animal cells. A unique feature of late endosomes, which are part of the pathway that leads to lysosomes, is that they contain a complex system of poorly characterized internal membranes in their lumen. These endosomes are therefore known as multivesicular or multilamellar organelles^{1,2}. Some proteins distribute preferentially within these internal membranes, whereas others are exclusively localized to the organelle's limiting membrane³. The composition and function of this membrane system are poorly understood. Here we show that these internal membranes contain large amounts of a unique lipid, and thus form specialized domains within endosomes. These specialized domains are involved in sorting the multifunctional receptor⁴ for insulin-like growth factor 2 and ligands bearing mannose-6-phosphate, in particular lysosomal enzymes. We also show that this unique lipid is a specific antigen for human antibodies associated with the antiphospholipid syndrome^{5,6}. These antibodies may act intracellularly by altering the protein-sorting functions of endosomes.

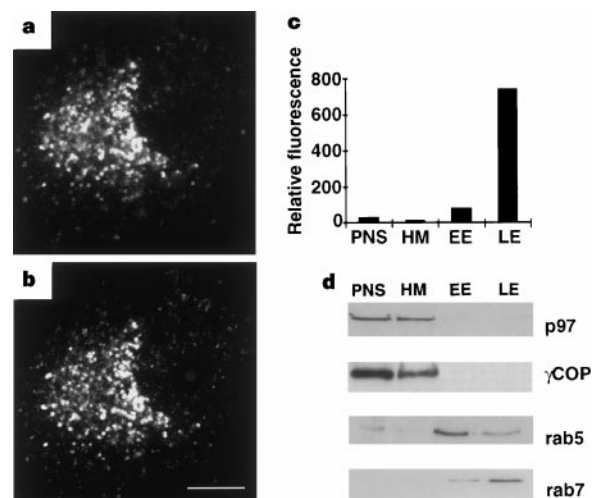


Figure 1 The 6C4 antigen is localized in late endosomes. **a, b**, BHK cells were fixed, permeabilized, and then double-labelled with 6C4 (**a**) and anti-rab7 (**b**) antibodies. Bar represents 10 μm. **c, d**, Subcellular fractions were analysed to establish the distribution of the following markers: **c**, 6C4 antigen using an ELISA assay; **d**, p97, γCOP, rab5 and rab7 by SDS-gel electrophoresis followed by western blotting. PNS, postnuclear supernatant; HM, heavy membranes; EE, early endosomes; LE, late endosomes.

We generated a monoclonal antibody (6C4), using endosomal membranes from BHK cells as antigens. The epitope recognized by 6C4 was absent from the cell surface (data not shown), and co-localized with the small GTPase rab7, a marker of late endosomes⁷ (Fig. 1a, b). We did not identify any protein recognized by the 6C4 antibody. However, using an enzyme-linked immunosorbent assay

(ELISA), we found that the total BHK lipid extract contained the immunoreactive material (Fig. 2d). The antigen was cellular, as no immunoreactivity was detected in the serum used to grow the cells; it was also sensitive to mild alkaline treatment (results not shown), and was therefore not a glycolipid. The antigen recognized by 6C4 co-purified with late endosomes containing rab7 (Fig. 1c, d), but not with p97 and γ COP, two proteins of the early protein-biosynthetic pathway⁸⁻¹⁰, or with the early-endosomal marker rab5¹¹. 6C4 therefore recognizes a late-endosomal lipid.

Late-endosomal fractions contained high amounts of triglycerides and cholesterol esters (Fig. 2a), and a limited subset of phospholipids (Fig. 2b), including a unique, highly hydrophobic, acidic phospholipid, which accounted for ~15% of the total phospholipids in the fractions (Table 1). This lipid migrated at the known position of lysobisphosphatidic acid (LBPA), and its deacylated form corresponded with the expected glycerolphosphorylglycerol backbone of LBPA (Fig. 2c). LBPA is not easily degraded by phospholipases¹², is structurally related to cardiolipin, and has been reported to distribute to lysosomes¹³. Purified LBPA was highly immunoreactive towards 6C4 in an ELISA assay (Fig. 2d); this immunoreactivity was confirmed by immunoblotting after chromatography (Fig. 2e). By dot blotting, 6C4 recognized only LBPA and none of the other lipids tested (Fig. 2f).

The precise subcellular distribution of LBPA was established by immunogold labelling of cryosections. As expected (Fig. 1), 6C4 labelled typical late-endosomal structures (Fig. 3b), which contained lgp120, an abundant membrane protein of BHK late

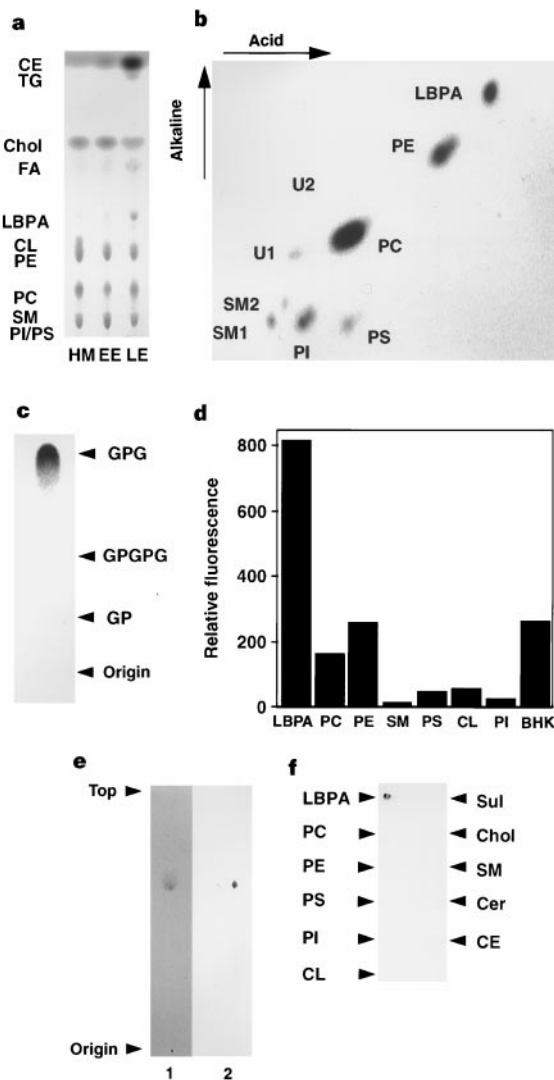


Figure 2 LBPA is enriched in late-endosome fractions and is the 6C4 antigen. **a**, Total lipid extracts (10-nmol phospholipid) from HM, EE or LE fractions (prepared as in Fig. 1c-d) were separated by TLC. CE, cholesterol ester; TG, triglyceride; Chol, cholesterol; FA, fatty acid; LBPA, lysobisphosphatidic acid; CL, cardiolipin; PE, phosphatidylethanolamine; PC, phosphatidylcholine; SM, sphingomyelin; PI, phosphatidylinositol; PS, phosphatidylserine. **b**, ³²P-labelled phospholipids of the LE fraction were extracted and separated by two-dimensional TLC. U1 and U2, unidentified spots; other abbreviations as in **a**. Sphingomyelin migrates as two spots in this system. **c**, The water-soluble product generated by mild alkaline methanolysis of ³²P-labelled LBPA. GPG, glycerophosphorylglycerol; GPGPG, bis(glycerophosphoryl)glycerol; GP, glycerophosphate. **d**, Microtitre wells were coated with 1.5 nmol of each lipid indicated and incubated with 6C4. Amounts of bound 6C4 were then quantified using a fluorescent-detection system (relative fluorescence signals are indicated). BHK, total lipid from BHK cells; other abbreviations as in **a**. **e**, Purified LBPA was developed on silica-gel 60 HPTLC with chloroform/methanol/32% ammonia (65:35:5, v/v). Lane 1 shows LBPA itself detected by phosphate staining, and lane 2 shows the same lane immunoblotted with 6C4 and analysed by enhanced chemiluminescence (ECL). **f**, 1 nmol of each lipid shown was spotted on an HPTLC plate, and then blotted with the 6C4 antibody. Sul, sulfatides; Cer, cerebroside; other abbreviations as in **a**.

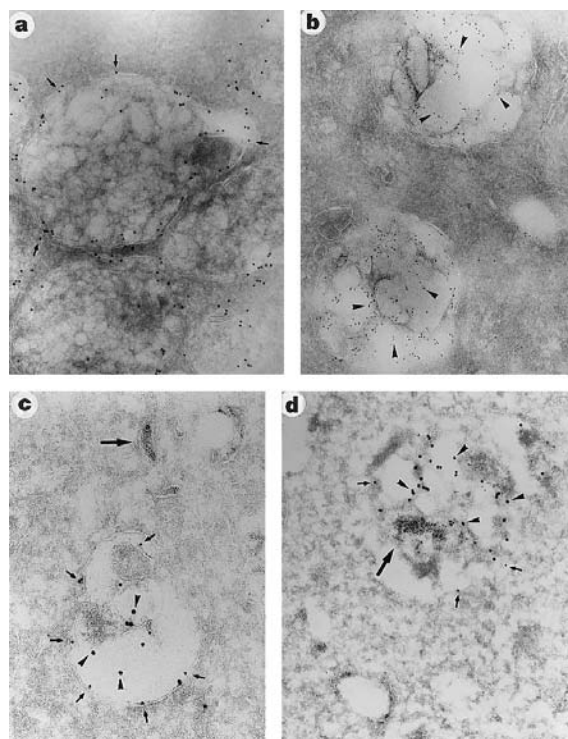


Figure 3 Electron microscopy. **a-b**, Cryosections were labelled with the 4A1 antibody against lgp120 (**a**, arrows) or the 6C4 antibody against LBPA (**b**, arrowheads) followed by rabbit antibodies against mouse IgG and then 10-nm (**a**) or 5-nm (**b**) protein-A gold. **c, d**, Early (**c**) or late (**d**) endosomes were labelled with 5-nm BSA-gold, which had been pre-internalized for 10 or 60 min at 37°C (thick arrows). Cryosections were double-labelled with 6C4 (arrowheads) and 15-nm protein-A gold followed by 4A1 (small arrows) and 10-nm protein-A gold. Micrographs were printed lightly to reveal BSA-gold particles (thick arrows). Labelling efficiency is reduced as compared with **a, b**, because the linker rabbit antibodies against mouse IgG were omitted for double-labelling.

endosomes¹⁴. Bovine serum albumin (BSA)-gold that had been endocytosed for 60 min, but not for 10 min, also labelled 6C4-positive late endosomes (Fig. 3c, d). However, 6C4-immunogold particles were only associated with the complex system of internal membranes; these membranes are characteristic of late endosomes in animal cells. The absence of immunoreactivity on the surface of the limiting endosomal membrane was confirmed biochemically by

comparing intact and disrupted late endosomes (results not shown). In contrast, Igp120 was found almost exclusively on the limiting membranes of the same endosomes (Fig. 3a, d)^{3,14}. The composition of these internal endosomal membranes is therefore different from that of the limiting endosomal membrane, indicating that the internal membranes form specialized microdomains within endosomes.

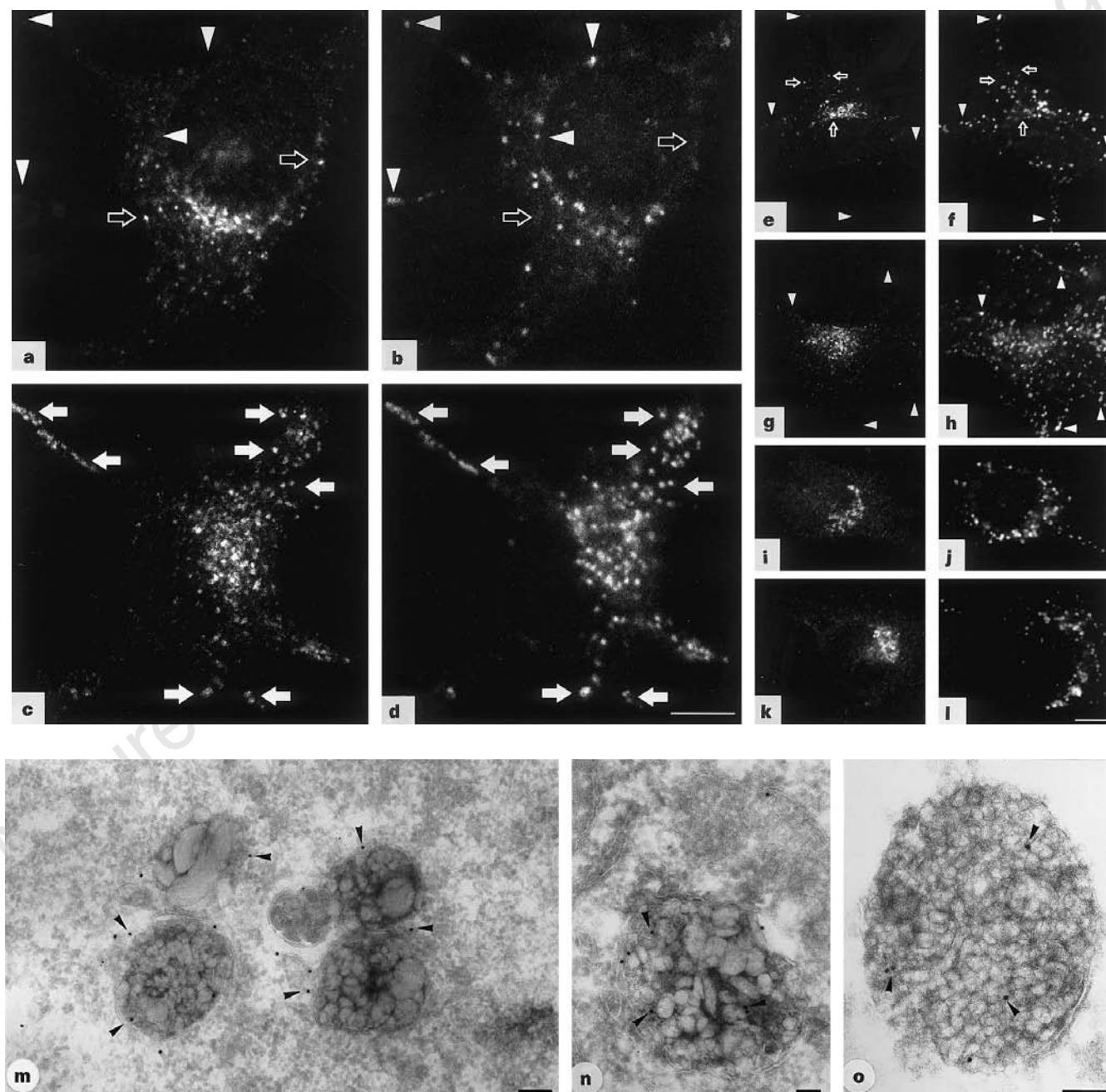


Figure 4 The ingested anti-LBPA antibody causes IGF2/MPR redistribution and alter late endosome ultrastructure. **a, b**, Steady-state distribution of IGF2/MPR (**a**) and LBPA (**b**) by indirect immunofluorescence microscopy. Arrows indicate IGF2/MPR-positive but LBPA-negative TGN elements; arrowheads indicate LBPA-positive but IGF2/MPR-negative late endosomes. **c, d**, Cells were incubated for 24 h with 50 $\mu\text{g ml}^{-1}$ 6C4 (anti-LBPA) to allow endocytosis of the antibody to occur. IGF2/MPR (**c**) shows extensive co-localization (arrows) with endocytosed 6C4 antibodies (**d**). **e, f**, Steady-state distribution of IGF2/MPR (**e**) and Igp120 (**f**). Arrows indicate IGF2/MPR-positive but Igp120-negative TGN elements; arrowheads indicate Igp120-positive but IGF2/MPR-negative late endosomes. **g, h**, as

for **c, d**, but 6C4 was replaced by 200 $\mu\text{g ml}^{-1}$ 4A1. Arrowheads indicate vesicles containing endocytosed antibodies against Igp120 (**h**) but not IGF2/MPR (**g, i, j**). Steady-state distribution of TGN38 (**i**) and LBPA (**j**). **k, l**, Antibodies against LBPA (**l**), endocytosed as in **c, d**, do not cause redistribution of TGN38 (**k**). **m-o**, Cells were pre-incubated for 24 h with 6C4, as in **c, d**. Cryosections were prepared and labelled with 4A1 (anti-Igp120, **m**), or rabbit antibodies against mouse IgG (to detect internalized 6C4, **n**) or anti-IGF2/MPR antibodies (**o**) followed by 15-nm protein-A gold. Micrographs were printed lightly (as in Fig. 3c-d) to reveal gold particles. Bars in **a-l** represent 10 μm and in **m-o** represent 100 nm.

To investigate the possible role of LBPA-rich membrane domains, 6C4 was internalized by fluid-phase endocytosis from the medium. As expected, the endocytosed antibody accumulated in late endosomes after binding to its antigen (Fig. 4d). In contrast, a non-relevant antibody could not be accumulated intracellularly, presumably because it was transported to lysosomes and then degraded (results not shown). One of the important functions of late endosomes is the sorting of the multifunctional receptor (IGF2/MPR)

Table 1 Phospholipid composition of the fractions

	HM	EE	LE
PC	52.9 ± 1.0	47.4 ± 0.3	49.2 ± 1.4
PE	18.2 ± 3.6	23.2 ± 1.8	18.5 ± 0.4
SM	5.3 ± 0.1	9.0 ± 0.3	3.2 ± 0.3
PS	5.9 ± 0.2	8.5 ± 0.2	3.9 ± 0.5
CL	3.8 ± 0.6	0.2 ± 0.2	0.2 ± 0.1
PI	8.6 ± 0.1	8.0 ± 1.0	7.0 ± 0.4
LBPA	0.8 ± 0.4	1.3 ± 0.4	13.8 ± 0.8
U1	3.3 ± 1.8	1.4 ± 0.5	2.4 ± 0.0
U2	1.2 ± 0.2	1.0 ± 0.1	2.0 ± 0.3

Cells were labelled with ^{32}P for 20 h before the experiments, homogenized, and then fractionated by flotation on a step sucrose gradient. Lipids from each fraction were extracted and then separated (see Fig. 2b). Radioactivity for each spot was quantified, and normalized to 100% for each fraction. Data represent the mean of two independent experiments. CL, cardiolipin; EE, early endosomes; HM, heavy membranes; LE, late endosomes; PC, phosphatidylcholine; PE, phosphatidylethanolamine; PI, phosphatidylinositol; PS, phosphatidylserine; SM, sphingomyelin; U1 and U2, unidentified spots.

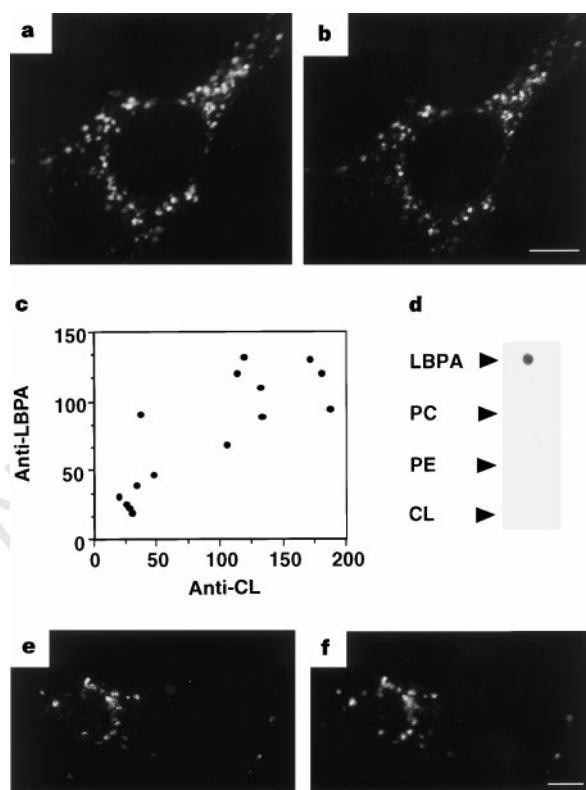


Figure 5 Sera from patients with antiphospholipid syndrome recognize LBPA. **a**, **b**, Cells were double-labelled with 6C4 (**a**) and aPL antiserum diluted 1:100 (**b**) and analysed by immunofluorescence. Three sera with a high anti-cardiolipin titer were tested (see **c**), and all labelled late endosomes containing LBPA. Bar represents 10 μm . **c**, Purified CL (cardiolipin) and LBPA were separately bound to microtitre wells, and reacted with 1:100 dilutions of different human sera. The anti-LBPA activity of each serum, quantified by ELISA, is plotted as a function of its anti-CL activity. **d**, The indicated lipids were spotted onto HPTLC plates and blotted with a 10-fold dilution of the serum used in **b**. Abbreviations as in Fig. 2. **e–f**, Cells were incubated for 13 h with 10% of the human serum used in **b** and processed as in Fig. 4c–d. IGF2/MPR (**e**) then co-localized with LBPA (**f**).

for mannose-6-phosphate-bearing ligands and insulin-like growth factor 2 (ref. 4). IGF2/MPR delivers newly synthesized lysosomal enzymes, which carry the mannose-6-phosphate signal, from the *trans*-Golgi network (TGN) to late endosomes, and must recycle back to the TGN for re-utilization. In the BHK cells, IGF2/MPR is predominantly localized to the TGN at steady state, and does not co-localize with LBPA to any significant extent (Fig. 4a, b). However, after treatment with 6C4, but not with control antibodies (results not shown), IGF2/MPR distribution was markedly changed. Most of the receptor was then found in late endosomes, and co-localized with LBPA (Fig. 4c, d; see also Fig. 4n, o) in >70% of the 6C4-treated cells, whereas co-localization occurred in <5% of untreated cells.

We performed the same experiment with anti-lgp120 antibody; lgp120 localizes to the limiting membrane of late endosomes. The antibody also accumulated in the cells (Fig. 4h), as would be predicted because the epitope is luminal, but IGF2/MPR trafficking was unaffected, even when using high antibody concentrations (Fig. 4j). Moreover, the distribution of TGN38, a protein which also cycles rapidly between TGN and endosomes and localizes to the TGN at steady state¹⁵ (Fig. 4i, j), was unaffected after 6C4 treatment (Fig. 4k, l). Ingested antibodies against LBPA therefore prevent proper IGF2/MPR trafficking in late endosomes in a highly selective manner, and thus interfere with late-endosome-sorting functions.

Endocytosed antibodies did not cause general perturbations of late endosomes/lysosomes, as the acidification properties of these cellular compartments were unaffected (results not shown). However, the treatment caused the appearance of a population of abnormal, electron-dense late endosomes, filled with densely packed membranes (Fig. 4m–o); such endosomes were rarely seen in control cells (see Fig. 3c, d). These abnormal late endosomes were lgp120-positive (Fig. 4m), and contained both endocytosed 6C4 antibodies (Fig. 4n) and IGF2/MPR (Fig. 4o), as expected (Fig. 4c, d). In contrast, uptake of anti-lgp120 antibodies (see Fig. 4h, j) did not alter endosome ultrastructure, nor did they cause lgp120 redistribution to internal membranes (results not shown). The striking disorganization of the inner-membrane system caused by anti-LBPA antibodies probably prevents proper segregation of IGF2/MPR molecules, thus causing them to be trapped within late endosomes. Indeed, IGF2/MPR, in contrast to lgp120, seems to segregate preferentially to LBPA-rich internal domains (Fig. 4o)³, although the mechanism that regulates this selective incorporation is unknown. However, our data suggest that anti-LBPA antibodies selectively alter the organization and dynamics of the complex system of internal membranes within late endosomes.

Human antiphospholipid antibodies are associated with a well described, yet poorly understood, antiphospholipid syndrome^{5,6} that is partly characterized by increased risk of thrombosis, recurrent fetal loss and thrombocytopenia¹⁶. A current view is that various antiphospholipid antibodies exist, some being directed against phospholipids alone, others being directed against protein–phospholipid complexes, and a third group being directed against proteins such as β 2-glycoprotein I (ref. 17). No antiphospholipid antigen had been localized at the cellular or subcellular level until now, and little is known about how these antibodies act. Antibodies of the antiphospholipid family that bind pure lipids in solid-phase assays are generally referred to as anti-cardiolipin, as cardiolipin is the most commonly used antigen, although these antibodies also bind phosphatidylserine and other anionic phospholipids to varying extents⁵.

As shown in Fig. 5a, b, human sera from patients with antiphospholipid syndrome, but not from healthy donors (results not shown), labelled late endosomes containing LBPA, but did not label the plasma membrane (which contains phosphatidylserine) or mitochondria (which contain cardiolipin). A comparative ELISA assay showed that these sera exhibited a strong immunoreactivity

against LBPA that correlated well with their established immunoreactivity against cardiolipin (Fig. 5c). Recent studies suggest that cardiolipin becomes oxidized when immobilized onto plastic in ELISA reactions, and that oxidation products, rather than cardiolipin itself, are recognized by patients' antisera¹⁸. We therefore tested the reactivity of human sera after spotting lipids directly onto high-performance thin-layer chromatography (HPTLC) plates, and found that only LBPA was antigenic (Fig. 5d). Finally, after incubation of living cells with serum from a patient with antiphospholipid syndrome, endocytosed antibodies accumulated in late endosomes containing LBPA (results not shown). Concomitantly, IGF2/MPR was redistributed, as observed after 6C4 uptake (Fig. 4c, d), and co-localized with LBPA in late endosomes (Fig. 5e–f); human sera from healthy donors had no effects (results not shown).

We show, therefore, that inner membranes of late endosomes are highly specialized and contain high amounts of the unconventional lipid LBPA, and that these LBPA-rich membranes play an important role in proper sorting and trafficking of IGF2/MPR. Our results indicate that LBPA is one of the physiological antigens that is recognized by sera from patients with antiphospholipid syndrome, and suggest that these antibodies exert some pathological effects intracellularly by altering endosomal sorting and/or trafficking of IGF2/MPR. As IGF2/MPR is multifunctional, changes in its transport cycle may have multiple effects, including in vesicular traffic¹⁹, lysosome biogenesis⁴, cell growth²⁰, and angiogenesis²¹. The antiphospholipid syndrome, particularly fetal loss, may be at least partly related to the role of IGF2/MPR in endothelial cell migration and neovascularization^{21,22}. □

Methods

Cells and reagents. Baby hamster kidney (BHK-21) cells were grown and maintained as described²³. Where indicated, cells were incubated for 24 h with medium supplemented either with hybridoma supernatant containing 50 $\mu\text{g ml}^{-1}$ 6C4 or 200 $\mu\text{g ml}^{-1}$ 4A1 monoclonal antibody, or with 10% human serum. The 6C4 monoclonal antibody (IgG) was generated after intrasplenic injection of immuno-isolated endosomes²³. The 4A1 monoclonal antibody (IgG) has been described¹⁴. Antibodies against rab5 or rab7 were raised against carboxy-terminal peptides⁷. Polyclonal antibodies against γCOP , IGF2/MPR and TGN38 were gifts from F. Wieland (Ruprecht Karl University, Heidelberg), B. Hoflack (Institut Pasteur, Lille), and G. Banting (University of Bristol), respectively. The anti-p97 monoclonal antibody was a gift from J. Peters (Harvard University, Boston). Sera from patients with antiphospholipid syndrome were from the Haemostasis Unit, University Hospital of Geneva. LBPA was purified from BHK lipid extracts by preparative thin-layer chromatography after silica gel and DEAE column chromatography; the purified lipid migrated as a single spot on HPTLC plates, was resistant to phospholipase A₂ (ref. 12), and did not react with ninhydrin, as expected. Egg-yolk phosphatidylcholine, egg phosphatidylethanolamine, brain phosphatidylserine, bovine-liver phosphatidylinositol, bovine-heart cardiolipin, brain sphingomyelin and brain cerebrosides were from Avanti polar lipids. Bovine-brain sulfatides, cholesterol and cholesterol oleate were from Sigma. Tetraoleoyl bisphosphatidic acid was from Doosan Serdary.

Lipid analysis. Subcellular fractions were prepared using a step-flotation gradient^{11,14}. Lipids were extracted and separated on silica-gel 60 HPTLC plates (Merck) by a two-step migration in the same direction, using chloroform/methanol/32% ammonia (65:35:5, v/v), then hexane/diethylether/acetic acid (16:4:2, v/v). Lipids were visualized after charring with cupric acetate. ³²P_i-labelled lipids were extracted and then separated by two-dimensional chromatography²⁴, detected by autoradiography, and quantified using the Molecular Imager System (BioRad). Mild alkaline methanolysis of LBPA was performed as described²⁵. The reference compounds were prepared from phosphatidic acid (glycerophosphate), cardiolipin (bis(glycerophosphoryl)-glycerol) and bisphosphatidic acid (glycerylphosphorylglycerol).

Immunodetection. An ELISA assay of lipids was performed as described²⁶; for human sera, microtitre wells were blocked with 10% FCS instead of 3% BSA (ref. 27). TLC plates were blotted with antibodies using a modified version of the method described in ref. 28. After treatment with polyisobutylmethacrylate

for 5 s, HPTLC plates were blocked for 2 h either with 20-mM HEPES at pH 7.4 and 150-mM NaCl (HEPES–NaCl) containing 3% BSA (for 6C4), or with 10% FCS in HEPES–NaCl (for human sera). After washing, the plates were incubated for 2 h with 10 $\mu\text{g ml}^{-1}$ 6C4 in HEPES–NaCl containing 1% BSA, or with 10× diluted human aPL antisera in HEPES–NaCl containing 10% FCS. The plates were reincubated with peroxidase-conjugated sheep antibodies against mouse IgG for 1.5 h. Bound antibodies were detected by chemiluminescence using the ECL reagent (Amersham).

Microscopy. Immunofluorescence was performed as described²⁹. To label early or late endosomes for electron microscopy, BHK cells were incubated for 10 or for 60 min, respectively, at 37 °C in the presence of 5-nm BSA–gold. Cells were then processed for immunogold labelling³⁰.

Received 11 September; accepted 5 December 1997.

- Mellman, I. Endocytosis and molecular sorting. *Annu. Rev. Cell Dev. Biol.* **12**, 575–626 (1996).
- Gruenberg, J. & Maxfield, F. Membrane transport in the endocytic pathway. *Curr. Opin. Cell Biol.* **7**, 552–563 (1995).
- Griffiths, G., Hoflack, B., Simons, K., Mellman, I. & Kornfeld, S. The mannose-6-phosphate receptor and the biogenesis of lysosomes. *Cell* **52**, 329–341 (1988).
- Kornfeld, S. Structure and function of the mannose-6-phosphate/insulin-like growth factor II receptors. *Annu. Rev. Biochem.* **61**, 307–330 (1992).
- McNeil, H. P., Chesterman, C. N. & Krilis, S. A. Immunology and clinical importance of antiphospholipid antibodies. *Adv. Immunol.* **49**, 193–280 (1991).
- Asherson, R. A., Cervera, R., Piette, J.-C. & Shoenfeld, Y. (eds) The antiphospholipid syndrome 1–339 (CRC Boca Raton, New York, London, Tokyo, 1996).
- Chavrier, P., Parton, R. G., Hauri, H. P., Simons, K. & Zerial, M. Localisation of low molecular weight GTP binding proteins to exocytic and endocytic compartments. *Cell* **62**, 317–329 (1990).
- Acharya, U. et al. The formation of Golgi stacks from vesiculated Golgi membranes requires two distinct fusion events. *Cell* **82**, 895–904 (1995).
- Rabouille, C., Levine, T. P., Peters, J.-M. & Warren, G. An NSF-like ATPase, p97, and NSF mediate cis/trans transport from mitotic Golgi fragments. *Cell* **82**, 905–914 (1995).
- Cosson, P. & Letourneur, F. Coatamer (COP1)-coated vesicles: role in intracellular transport and protein sorting. *Curr. Opin. Cell Biol.* **9**, 484–487 (1997).
- Gorvel, J. P., Chavrier, P., Zerial, M. & Gruenberg, J. Rab 5 controls early endosome fusion *in vitro*. *Cell* **64**, 915–925 (1991).
- Brotherus, J., Renkonen, O., Herrmann, J. & Fisher, W. Novel stereochemical configuration in lysobisphosphatidic acid of cultured BHK cells. *Chem. Phys. Lipids* **13**, 178–182 (1974).
- Wherrett, J. R. & Huterer, S. Enrichment of bis(monoacylglycerol)phosphate in lysosomes from rat liver. *J. Biol. Chem.* **247**, 4114–4120 (1972).
- Aniento, F., Emans, N., Griffiths, G. & Gruenberg, J. Cytoplasmic dynein-dependent vesicular transport from early to late endosomes. *J. Cell Biol.* **123**, 1373–1388 (1993).
- Reaves, B., Horn, M. & Banting, G. TGN38/41 recycles between the cell surface and the TGN: brefeldin A affects its rate of return to the TGN. *Mol. Biol. Cell* **4**, 93–105 (1993).
- Khamashta, M. A., Gharavi, A. E. & Wilson, W. A. (eds) New Orleans 7th Int. Symp. on Antiphospholipid Antibodies: Proceedings and Abstracts. *Lupus* **5**, 343–558 (1996).
- Alarcon-Segovia, D. & Cabral, A. R. The antiphospholipid/cofactor syndromes. *J. Rheumatol.* **23**, 1319–1322 (1996).
- Hörkkö, S. et al. Antiphospholipid antibodies are directed against epitopes of oxidized phospholipids. *J. Clin. Invest.* **98**, 815–825 (1996).
- Ludwig, T., Le Borgne, R. & Hoflack, B. Roles for mannose-6-phosphate receptors in lysosomal sorting, IGF-II binding and clathrin coat assembly. *Trends Cell Biol.* **5**, 202–205 (1995).
- Wang, Z. Q., Fung, M. R., Barlow, D. P. & Wagner, E. F. Regulation of embryonic growth and lysosomal targeting by the imprinted Igf2/Mpr gene. *Nature* **372**, 464–467 (1994).
- Volpert, O., Jackson, D., Bouck, N. & Linzer, D. I. The insulin-like growth factor II/mannose 6-phosphate receptor is required for proliferin-induced angiogenesis. *Endocrinology* **137**, 3871–3876 (1996).
- Groskopf, J. C., Syu, L. J., Saltiel, A. R. & Linzer, D. I. Proliferin induces endothelial cell chemotaxis through a G protein-coupled, mitogen-activated protein kinase-dependent pathway. *Endocrinology* **138**, 2835–2840 (1997).
- Gruenberg, J., Griffiths, G. & Howell, K. E. Characterisation of the early endosome and putative endocytic carrier vesicles *in vivo* and with an assay of vesicle fusion *in vitro*. *J. Cell Biol.* **108**, 1301–1316 (1989).
- Joutti, A. The stereoconfiguration of newly formed molecules of bis(monoacylglycerol)phosphate in BHK cells. *Biochim. Biophys. Acta* **575**, 10–15 (1979).
- Somerharju, P. & Renkonen, O. Glycerophospho-(N-acyl)-ethanolamine lipids in degenerating BHK cells. *Biochim. Biophys. Acta* **573**, 83–89 (1979).
- Umeda, M., Igarashi, K., Nam, K. S. & Inoue, K. Effective production of monoclonal antibodies against phosphatidylserine: stereo-specific recognition of phosphatidylserine by monoclonal antibody. *J. Immunol.* **143**, 2273–2279 (1989).
- Reber, G., Tremblat, C., Bernard, C., Mermillod, B. & de Moerloose, P. Anticardiolipin antibodies and thrombosis: buffer's influence on the detection and quantitation of anticardiolipin antibody measured by ELISA. *Thromb. Res.* **57**, 215–226 (1990).
- Karasawa, K., Satoh, N., Nakagawa, Y., Setaka, M. & Nojima, S. Specific binding of antibodies to platelet activating factor (PAF) as demonstrated by thin-layer chromatography/immunostaining. *Lipids* **26**, 1122–1125 (1991).
- Gu, E., Aniento, F., Parton, R. & Gruenberg, J. Functional dissection of COP-I subunits in the biogenesis of multivesicular endosomes. *J. Cell Biol.* **139**, 1183–1195 (1997).
- Griffiths, G., McDowell, A., Back, R. & Dubochet, J. On the preparation of cryosections for immunocytochemistry. *J. Ultrastruct. Res.* **89**, 65–78 (1984).

Acknowledgements. We thank M.-H. Beuchat for technical assistance; G. Griffiths for his help and support in the analysis of the monoclonal antibodies, H. Hardersen for preparing monoclonal antibodies; J. Deshusses for help in LBPA identification; M. Lindsay for help with electron microscopy; and G. van der Goot, U. Schiebler and F. Perez for reading the manuscript. This work was supported by grants from the Swiss National Science Foundation (to J.G.), the NHMRC of Australia (to R.G.P.), and the International Human Frontier Science Program (to J.G., R.G.P. and T.K.).

Correspondence and requests for materials should be addressed to J.G. (e-mail: jean.gruenberg@biochem.unige.ch).

Translocation of calmodulin to the nucleus supports CREB phosphorylation in hippocampal neurons

Karl Deisseroth*, E. Kevin Heist† & Richard W. Tsien*

* Department of Molecular and Cellular Physiology, Beckman Center for Molecular and Genetic Medicine, and †Department of Neurobiology, Stanford University School of Medicine, Stanford, California 94305-5426, USA

Activation of the transcription factor CREB is thought to be important in the formation of long-term memory in several animal species^{1–3}. The phosphorylation of a serine residue at position 133 of CREB is critical for activation of CREB⁴. This phosphorylation is rapid when driven by brief synaptic activity in hippocampal neurons⁵. It is initiated by a highly local, rise in calcium ion concentration⁵ near the cell membrane, but culminates in the activation of a specific calmodulin-dependent kinase known as CaMKIV (ref. 7), which is constitutively present in the neuronal nucleus^{7,8}. It is unclear how the signal is conveyed from

the synapse to the nucleus. We show here that brief bursts of activity cause a swift (~1 min) translocation of calmodulin from the cytoplasm to the nucleus, and that this translocation is important for the rapid phosphorylation of CREB. Certain Ca²⁺ entry systems (L-type Ca²⁺ channels and NMDA receptors) are able to cause mobilization of calmodulin, whereas others (N- and P/Q-type Ca²⁺ channels) are not. This translocation of calmodulin provides a form of cellular communication that combines the specificity of local Ca²⁺ signalling with the ability to produce action at a distance.

In hippocampal neurons subjected to synaptic stimulation or direct depolarization, Ca²⁺-dependent CREB phosphorylation occurs even when elevation in bulk [Ca²⁺]_i has been suppressed with internal EGTA⁵, suggesting that the signalling from synapse to nucleus does not involve the spread of free Ca²⁺ *per se*. We looked instead for translocation of a Ca²⁺ effector. Examination of an array of calmodulin (CaM) kinases^{9,10} in CA3/CA1 hippocampal pyramidal neurons, including CaMKI, CaMKIIα, CaMKIIβ, CaMKIV and CaMKKα, failed to reveal any striking redistribution upon strong direct depolarization of the neurons (for 180 s in 90 mM K⁺) (Fig. 1a). Although calmodulin itself is largely immobile in quiescent cells¹¹, changes in its localization to various cellular compartments, including to the nucleus, have been described^{11–13}. We therefore investigated whether the CaM translocation could be rapid enough to account for signalling on a timescale of minutes.

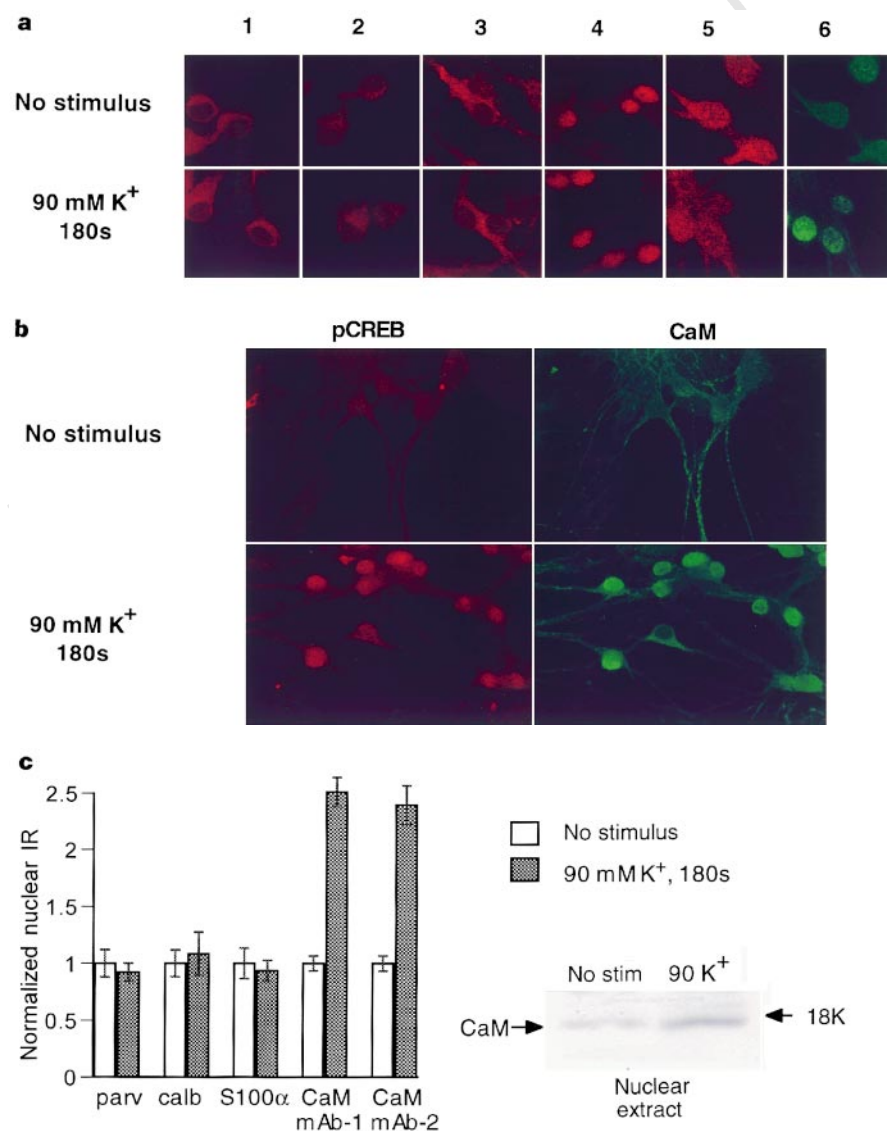


Figure 1 Rapid activity-dependent translocation of calmodulin to the nucleus in hippocampal CA3/CA1 pyramidal neurons. **a**, Confocal mid-nuclear sections showing CaMKI (1), CaMKIIα (2), CaMKIIβ (3), CaMKIV (4), CaMKKα (5), and CaM (6) immunofluorescence with or without stimulation. The same field of cells is shown in panels 5 and 6. **b**, Dendritic/nuclear confocal sections showing CaM translocation and CREB phosphorylation as a result of depolarization. **c**, Quantification of nuclear confocal data on CaM and other Ca²⁺-binding proteins. No activity-dependent redistribution to the nucleus was observed for parvalbumin, calbindin D28K, or S100α ($n > 50$ neurons; $P > 0.2$, error bars signify ± 1 s.e.m.). pCREB immunoreactivity was found to be increased in the same neurons, verifying that they had been successfully stimulated. In contrast, the 6D4 anti-CaM (mAb-2) revealed an increase in nuclear CaM, similar to results obtained with the UBI antibody (mAb-1) ($P < 0.001$ in both cases). Inset, nuclear extracts from stimulated cells showed increased CaM content (2.4-fold), as assessed by immunoblot.

We found a rapid, activity-dependent increase in CaM immunoreactivity (CaM-IR) in the nucleus by using a monoclonal antibody that recognizes CaM regardless of whether it is associated with Ca^{2+} (ref. 29) or a target kinase¹⁴ (Fig. 1a). Nuclear CaM was readily detectable without stimulation, but increased strongly after stimulation. The CaM-IR (Fig. 1a, mid-nuclear sections; Fig. 1b, right, dendritic/nuclear sections) appeared to be closely correlated with nuclear CREB phosphorylation (Fig. 1b, left). Similar results were obtained within 42 s of applying briefer stimuli (18 s of 50 Hz electrical field stimulation, Fig. 2; or 18 s of 90 mM K^+ , not shown). Comparable increases in nuclear CaM-IR were found with a second monoclonal antibody (mAb-2; Fig. 1c). In contrast, no redistribution was seen for other Ca^{2+} -binding proteins, parvalbumin, calbindin D28K, and S100 α , in neurons expressing these markers (Fig. 1c). A parallel activity-dependent reduction in cytosolic CaM-IR was measured at the apical dendrite (52% $\pm$ 9% of control, $P < 0.01$). CaM translocation slowly reversed on a time-scale of tens of minutes after stimulation (Fig. 4c). We confirmed the increase of nuclear CaM by subcellular fractionation. Depolarization (180 s of 90 mM K^+) caused a sizeable increase (2.4-fold) in nuclear CaM (Fig. 1c, inset). In a third approach, we tracked CaM translocation using a merocyanine-labelled calmodulin, MeroCaM¹⁵, introduced into individual neurons by patch pipette. Rapid (~ 1 min) elevations in nuclear MeroCaM were observed upon depolarization, whereas opposite changes in MeroCaM fluorescence were observed in cytoplasm, consistent with a net translocation (not shown); in addition, ratiometric imaging (F_{605}/F_{530}) demonstrated a rapid (~ 1 min) increase in the nuclear ratio of Ca^{2+} /MeroCaM to apo-MeroCaM, indicating that the proportion of nuclear CaM in the Ca^{2+} -bound state was also increased by membrane depolarization.

Brief bursts of electrical stimulation (50 Hz for 18 s) also gave rise

to a marked translocation of CaM to the nucleus (Fig. 2a, b), which was sharply diminished by the L-type Ca^{2+} -channel antagonist nimodipine (10 μM). The stimulus-induced rise in nuclear CaM-IR was also blunted by the NMDA-receptor antagonist D-AP5 (50 μM) and was eliminated completely when D-AP5 and nimodipine were applied together, whereas maximal blockade of N-type Ca^{2+} channels by ω -CTx GVIA (1 μM) barely reduced CaM. This latter effect could be due to the inhibition of presynaptic N-type channels contributing to synaptic glutamate release. To focus on the postsynaptic N-type channels, we bypassed the presynaptic terminal by imposing direct depolarization (90 mM K^+) in the presence of AP5 (Fig. 2c, d). In this case, ω -CTx GVIA did not reduce CaM translocation. Likewise, blockade of postsynaptic P/Q-type channels with saturating ω -AgaIVA (1 μM) was ineffective, whereas nimodipine blocked CaM translocation (Fig. 2d). We compared the contributions of the various channel types^{6,16,17} to depolarization-induced rises in somatic intracellular calcium ion concentration ($[\text{Ca}^{2+}]_i$) (Fig. 2d) under the conditions used to study CaM translocation (Fig. 2c). The contribution of L-type channels was significantly less than that of P/Q-type channels at every time point, and less than that of N-type channels at the time of peak $[\text{Ca}^{2+}]_i$, although the L-type channels contributed more to the plateau of $[\text{Ca}^{2+}]_i$ than did the N-type channels. Thus, CaM translocation cannot result simply from a bulk increase in $[\text{Ca}^{2+}]_i$, suggesting that coupling could be more specific between L-type channel activity and CaM or CaM-binding proteins.

If CaM translocation is necessary for signalling to the nucleus, the requirements for Ca^{2+} entry that apply to initiation of CaM translocation should also extend to the downstream nuclear event. Although L-type Ca^{2+} channels have long been implicated in gene expression^{18,19}, the importance of N- or P/Q-type channels has not

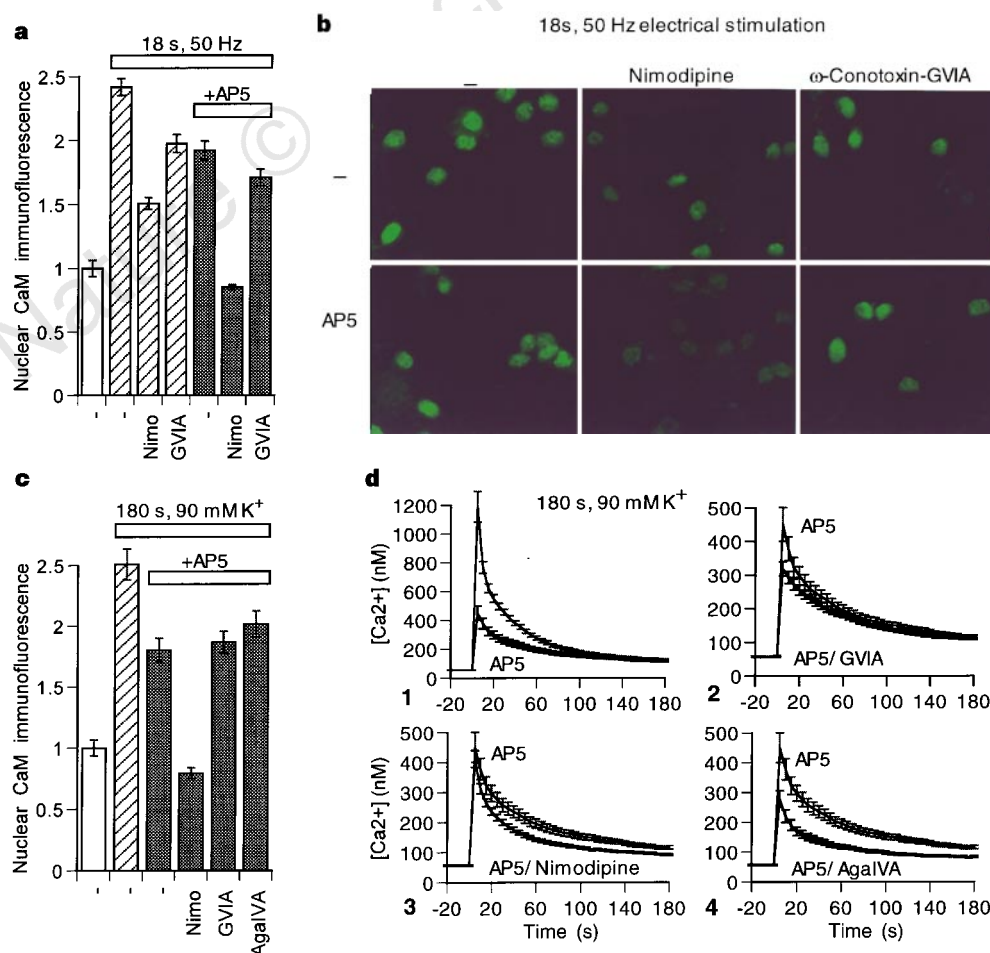


Figure 2 Selective control of activity-dependent CaM translocation by NMDA receptors and L-type Ca^{2+} channels. **a**, Effects of Ca^{2+} influx inhibitors on CaM translocation to the nucleus following brief electrical stimulation. All antagonist effects were highly significant ($P < 0.001$, $n > 50$ cells for each condition). **b**, Representative confocal sections used for analysis in **a**. **c**, Effects of Ca^{2+} influx inhibitors on CaM translocation to the nucleus following stimulation with 90 mM K^+ (3 min). In the presence of D-AP5, the nimodipine effect was highly significant ($P < 0.001$), whereas the effects of GVIA or AgaIVA were not ($P > 0.2$). **d**, Intracellular free $[\text{Ca}^{2+}]_i$ at the cell body measured with Fura-2 during depolarization under the conditions of **c**. D-AP5 strongly reduced the 90 mM K^+ -induced Ca^{2+} transient (1). Reduction of the somatic Ca^{2+} signal in D-AP5 by the Ca^{2+} channel antagonists ω -conotoxin GVIA (2), nimodipine (3), or ω -AgaIVA (4). The effect of ω -conotoxin GVIA was greater than the effect of nimodipine over the first 15 s of stimulation, but not beyond this point. The effect of AgaIVA was significantly greater than the effect of nimodipine throughout ($P < 0.001$ at the beginning, $P < 0.05$ at the end).

been systematically examined. We found a pharmacological profile for CREB Ser 133 phosphorylation²⁰ (Fig. 3a, b) very similar to that of CaM mobilization (Fig. 2a–c). Blockade of P/Q- or N-type channels was completely ineffective in reducing CREB phosphorylation (Fig. 3b), despite their strong participation in triggering an increase in $[Ca^{2+}]_i$ (Fig. 2d). The privileged linkage between L-type Ca^{2+} entry and CaM translocation to the nucleus thus extends to the putative outcome of the translocation.

To test the importance of nuclear CaM in synaptically induced CREB phosphorylation, we compared the immunoreactivities of CaM-IR and Ser 133-pCREB-IR in individual nuclei across a large population of cells. There was a strong correlation between levels of nuclear CaM and nuclear pCREB after 50 Hz electrical stimulation (Fig. 3c, $r = 0.907$, $n = 201$, $P < 0.0001$; Fig. 1b). The relation was maintained under partial pharmacological inhibition (Fig. 3d): the greater the reduction in nuclear CaM, the greater the decrease in pCREB. Like the results of Ca^{2+} -channel blockade, our results support the idea that nuclear CaM could be limiting in rapid CaM-kinase-dependent CREB phosphorylation. In contrast, stimulation of cyclic-AMP signalling with forskolin increased pCREB, weakly at 5 min and strongly at 30 min, but with no clear correlation with changes in nuclear CaM (Figs 3d, 4d). Presumably, forskolin promotes CREB phosphorylation through translocation of protein kinase A to the nucleus^{21,22} and not through the CaM translocation and CaM kinase mechanism recruited by synaptic activity. The efficacy of CaM translocation as triggered by Ca^{2+} influx through L-type channels was demonstrated by using a direct activator of L-type channel activity, FPL 64176 (Fig. 3e). When applied for 3 min, this agent increased nuclear CaM-IR several-fold, again with strongly correlated increases in nuclear pCREB. Previous work has shown a substantial increase in nuclear calmodulin after 1 h of Bay K8644 treatment *in vivo*¹³.

To interfere with CaM import into the nucleus, we exploited

the extreme temperature-sensitivity¹² of this process (diffusive processes such as entry of the protein kinase A catalytic subunit into the nucleus are only slightly temperature-sensitive²²). Moderate chilling after a brief burst of electrical activity (Fig. 4a) essentially abolished rapid translocation of CaM, although a slow leakage of CaM into the nucleus persisted at long times after the stimulus. This indicates that the nuclear translocation of CaM is not simply diffusive but may be facilitated¹², perhaps involving the cytoskeleton. Like CaM translocation, activity-dependent CREB phosphorylation is also sharply reduced by chilling, giving only a late increase at the lowered temperature (Fig. 4b). In contrast to behaviour triggered by synaptic activity, forskolin-stimulated CREB phosphorylation is not slowed by chilling²² (Fig. 4d). No significant CaM translocation was observed with forskolin treatment, regardless of temperature (Fig. 4d). Reversal of CaM translocation reached baseline levels 2 h after stimulation (Fig. 4c).

To intercept CaM in the nucleus, we transfected hippocampal neurons²³ with a CaM-inhibitor peptide specifically targeted to the nucleus²⁴ (nCaMBP) (Fig. 5). The peptide's nuclear location was confirmed by staining with an nCaMBP-specific antibody²⁴ and counterstaining with DAPI (Fig. 5a). Depolarization with 20 mM K^+ produced a clear increase in pCREB that was abolished by nCaMBP (Fig. 5b, middle) but not in control transfected cells (Fig. 5b, left). When the strength of the depolarizing stimulus was increased, nCaMBP only partly reduced the increase in pCREB (Fig. 5b, c), consistent with the competitive mechanism of nCaMBP action²⁴ or perhaps with recruitment of an additional pathway by this very strong stimulus.

Our experiments have revealed a rapid, activity-dependent nuclear translocation of calmodulin and demonstrated its importance for neuronal CREB phosphorylation. The time course of CaM translocation is fast enough (~ 1 – 2 min) to account for the kinetics of nuclear CREB phosphorylation under similar conditions^{5,7}.

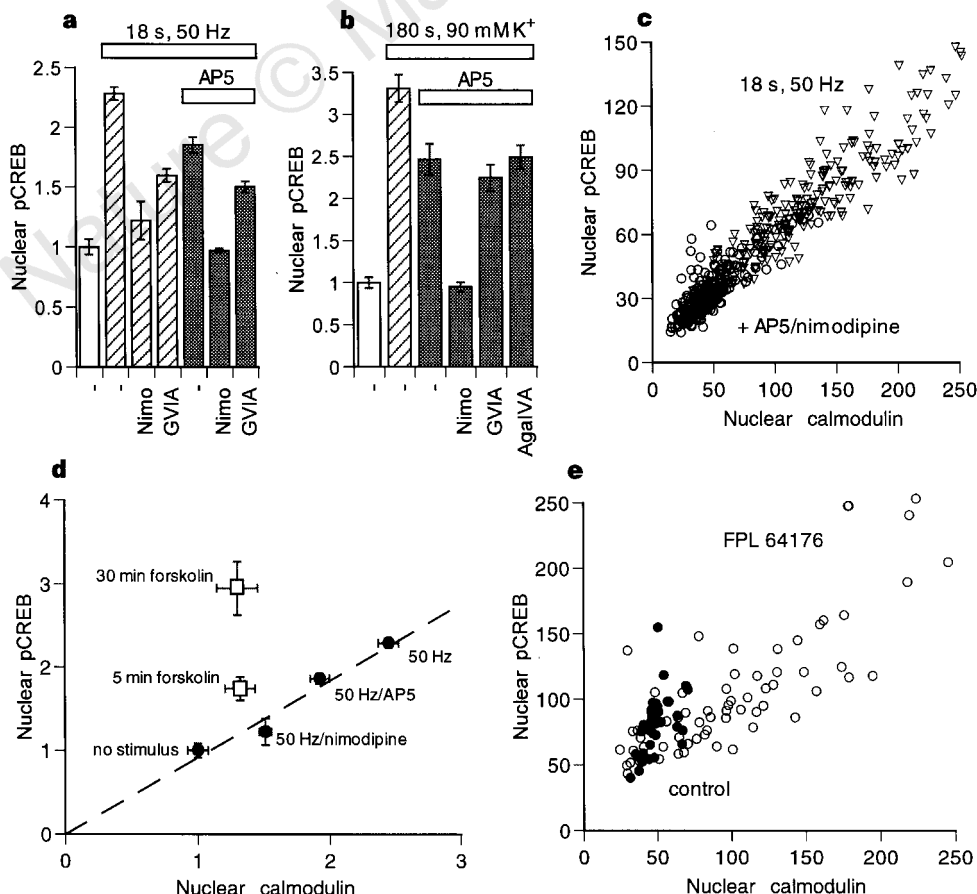


Figure 3 Close relationship between activity-dependent CaM translocation and CREB phosphorylation. **a**, Confocal quantification of CREB phosphorylation after synaptic stimulation at 50 Hz for 18 s. Experimental conditions and inhibitor concentrations as in Fig. 2a. All effects were highly significant relative to control ($P < 0.001$). **b**, Confocal quantification of CREB phosphorylation after 90 mM K^+ stimulation. The effects of α -AP5 alone and of nimodipine in α -AP5 were highly significant ($P < 0.001$), whereas the effects of ω -CTx-GvIA and ω -Aga-IVA in α -AP5 were not ($P > 0.2$). **c**, Single-cell correlation between CaM translocation and activity-dependent CREB phosphorylation, in response to 50 Hz electrical stimulation for 18 s (triangles, $n = 201$ cells, $r = 0.907$, $P < 0.0001$). Addition of α -AP5 and nimodipine (circles) reduced nuclear CaM and pCREB in parallel. **d**, Summary plot of mean values of CaM translocation and CREB phosphorylation under various stimulation conditions. **e**, FPL 64176 (5 μ M) was applied in external solution containing TTX, α -AP5, ω -conotoxin GvIA, ω -Aga-IVA, and CNQX (open symbols) and compared with sham controls (filled symbols).

Indeed, the arrival of CaM at the nucleus is critical for promoting CREB phosphorylation, as indicated by the insufficiency of basal nuclear CaM (Figs 2c, d, 3b), by the correlation between increased nuclear CaM and pCREB (Fig. 3), and by the effects on pCREB of blocking nuclear CaM import (Fig. 4) or of intercepting CaM in the nucleus (Fig. 5). Reinforcing the idea that signalling to the nucleus may originate from changes in local rather than in bulk $[Ca^{2+}]_i$ (refs 5, 25), even large increases in bulk $[Ca^{2+}]_i$ fail to promote CREB phosphorylation if not coupled with CaM translocation (Figs 2c, d, 3b). In fact, Ca^{2+} ions coming in may only need to reach apo-CaM near the plasma membrane to transmit the signal; molecules with a selective affinity for Ca^{2+} /CaM could be involved in CaM movement in the cell and would be expected to stabilize the Ca^{2+} -bound state. One nuclear target could be CaMKIV (refs 26, 27), which has been localized to the nucleus of hippocampal neurons and implicated in neuronal activity-dependent CREB phosphorylation⁷. CaMKIV can be phosphorylated and thereby activated by similarly CaM-dependent²⁸ kinase kinases. An isoform of CaMKK might translocate together with CaM, stabilizing Ca^{2+} -bound CaM and increas-

ing the activity of CaMKIV. The exceptionally fast kinetics of this pathway activated by bursts of synaptic activity may be well suited for communication of rapid information from synapse to nucleus. □

Methods

Electrophysiology. CA3/CA1 hippocampal neurons were cultured essentially as described⁵. Synaptic activity was controlled by field electrical stimulation⁵; unlike high- K^+ depolarization, action potentials provided by the field electrodes are not sufficient when postsynaptic receptors are blocked to give rise to CREB phosphorylation, probably because fast action potentials are too brief to open slowly-activating L-type channels sufficiently. Neurons were preincubated for 2–3 h before stimulation in 1 μ M tetrodotoxin (TTX) (Calbiochem); TTX was removed just before stimulation. Pharmacological agents were ω -agatoxin IVA and ω -conotoxin GVIA (Peptide Institute or Pfizer); nimodipine, D-AP5, FPL 64176, CNQX and forskolin (RBI).

Imaging. Imaging of Ca^{2+} and MeroCaM was done using a Videoscope ICCD camera. For Fura-2 Ca^{2+} imaging, data were taken from a measurement region encompassing the entire soma, including the nucleus. Neurons were fixed in

Figure 4 Requirement for nuclear CaM translocation in activity-dependent CREB phosphorylation. **a**, Post-stimulus chilling inhibits activity-dependent CaM translocation; this inhibition can be overcome at long times after the stimulus. **b**, Parallel inhibition of CREB phosphorylation under the same conditions. **c**, Time course of reversal of CaM translocation after cessation of brief electrical stimulation. Inset, time course of intracellular free $[Ca^{2+}]$ elevation in response to the stimulation (averaged trace, $n = 13$). **d**, cAMP-dependent CREB phosphorylation is not slowed by chilling.

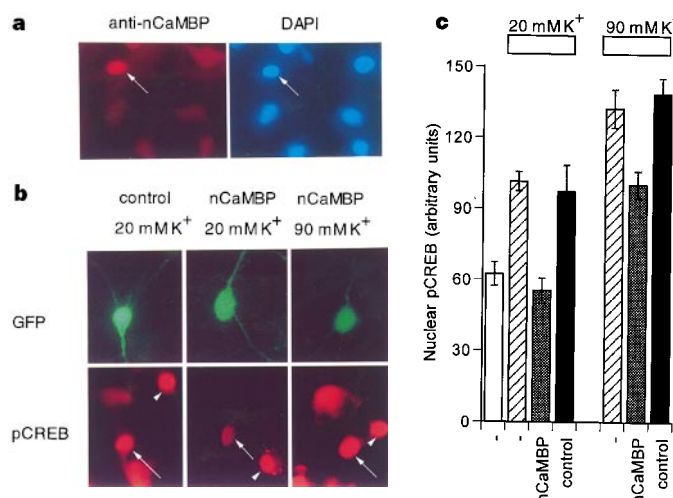
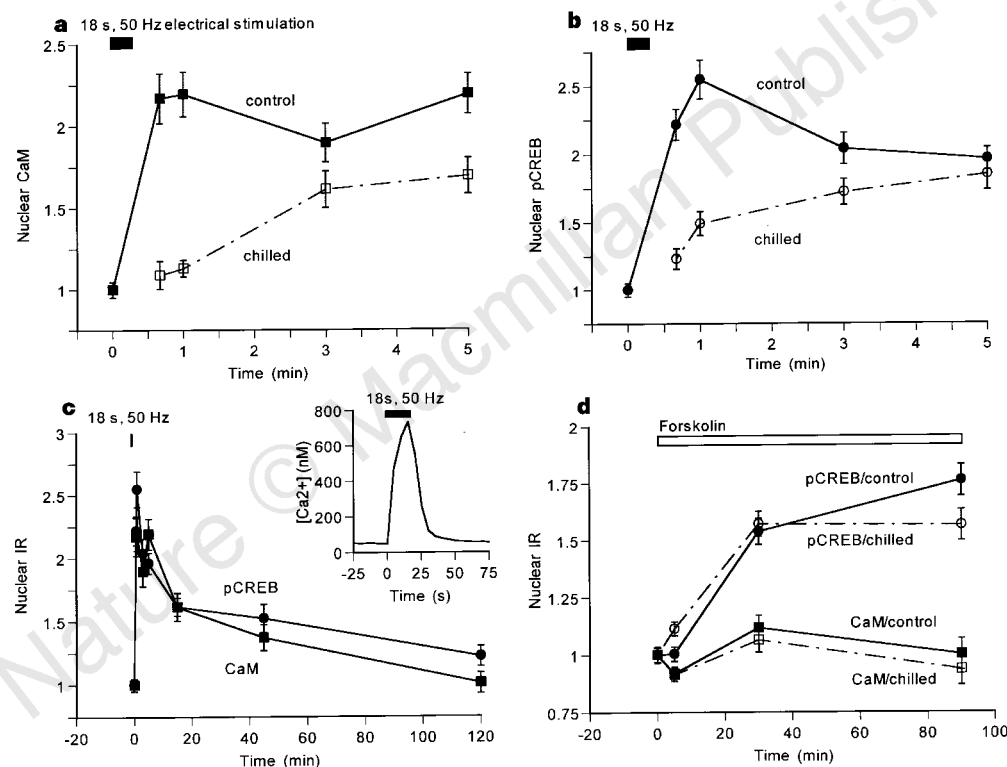


Figure 5 Causal link between nuclear CaM and activity-dependent CREB phosphorylation in hippocampal neurons. **a**, Nuclear localization of the transfected nuclear CaM binding protein (nCaMBP). DAPI stain (right) for DNA labels nuclei. **b**, Inhibition of activity-dependent (20 mM K^+) CREB phosphorylation by expression of nCaMBP (centre) but not in control transfection (left). Depolarization with 90 mM K^+ (right) was able largely to overcome the effect of nCaMBP. Arrows indicate transfected neurons identified by GFP fluorescence (GFP signal shown in upper row); arrowheads indicate non-transfected neurons nearby. Expression of nCaMBP did not affect neuronal morphology or total CREB levels (CREB immunofluorescence, $105.2 \pm 14.5\%$ of control, $n = 8$). **c**, nCaMBP abolished 20 mM K^+ -induced CREB phosphorylation ($P < 0.01$, $n = 9$); empty vector had no effect ($P > 0.2$, $n = 14$). As a separate control, we coexpressed β -galactosidase instead of nCaMBP; no effect was seen on CREB phosphorylation (pCREB-IR $101.4 \pm 11.2\%$ relative to control, $n = 7$). nCaMBP partially inhibited 90 mM K^+ -induced CREB phosphorylation ($P < 0.05$, $n = 16$), whereas transfection with empty vector did not ($P > 0.2$, $n = 13$).

ice-cold 4 mM EGTA/4% formaldehyde/PBS and permeabilized in 0.4% saponin or 0.1% Triton X-100 (fainter background staining was detected with saponin). The monoclonal anti-CaM antibodies (mAb-1 from UBI was used for all experiments except for Fig. 1c, for which mAb-2 from Sigma was also used) each recognize both Ca^{2+} -bound and Ca^{2+} -free forms of CaM^{29,30}; the UBI antibody has a slight preference for Ca^{2+} -free CaM and the Sigma antibody Ca^{2+} -bound CaM; as all the fixations were in 4 mM EGTA, the antibodies gave quantitatively indistinguishable results (Fig. 1c). Other primary antibodies used were anti-calbindin D28K, anti-parvalbumin and anti-S100 α (Sigma), anti-pCREB and anti-CREB (UBI), anti-CaMKI (gift from A. Nairn), anti-CaMKII α and anti-CaMKII β (Life Technologies), anti-CaMKIV and anti-CaMKK α (Santa Cruz Biotechnology), and anti-nCaMBP (gift from M. Kaetzel and J. Dedman). For chilling experiments, cells were bathed at 7.5 °C or 24 °C within 5 s of electrical stimulation. Confocal sections were mid-nuclear; DAPI (4,6-diamidino-2-phenylindole) was used to validate the morphological identification of nuclei.

Transfection. Calcium phosphate transfections of neurons²³ were carried out in 24-well plates using 4 μg DNA per well. Neurons were exposed to the precipitate for 2 h in serum-free OptiMem (GIBCO/BRL) and subsequently returned to normal culture medium for 48 h. In co-transfections, the green fluorescent protein (GFP) marker plasmid was put at a disadvantage to ensure that GFP-positive cells express the cotransfected plasmid, as verified by double-marker experiments. The E-GFP expression vector (pEGFP-C1) was from Clontech; other materials were gifts from J. Wang, M. Kaetzel and J. Dedman and have been described²⁴.

Subcellular fractionation. Fractionation was done essentially as described⁵: to hypotonic lysis buffer were added 2 mM CaCl_2 , 1 mM MgCl_2 , 1 mini-protease inhibitor tablet (Boehringer), and 2.5 mg ml^{-1} wheat germ agglutinin (WGA, Sigma; this WGA fraction exhibited some nonspecific binding to antibodies on western blot). Fractionation was carried out as near to 0 °C as possible. Protein concentrations were determined by the Bradford method, and 4 μg of each sample was loaded onto each lane for denaturing 15% SDS-PAGE and then transferred to nitrocellulose. The membrane was probed with anti-CaM monoclonal antibody (UBI) and detected by ECL.

Received 15 September; accepted 26 November 1997.

1. Dash, P. K., Hochner, B. & Kandel, E. R. Injection of the cAMP-responsive element into the nucleus of *Aplysia* sensory neurons blocks long-term facilitation. *Nature* **345**, 718–721 (1990).
2. Bourtchuladze, R. et al. Deficient long-term memory in mice with a targeted mutation of the cAMP-responsive element-binding protein. *Cell* **79**, 59–68 (1994).
3. Yin, J. C. et al. Induction of a dominant negative CREB transgene specifically blocks long-term memory in *Drosophila*. *Cell* **79**, 49–58 (1994).
4. Gonzalez, G. A. & Montminy, M. R. Cyclic AMP stimulates somatostatin gene transcription by phosphorylation of CREB at serine 133. *Cell* **59**, 675–680 (1989).
5. Deisseroth, K., Bito, H. & Tsien, R. W. Signaling from synapse to nucleus: postsynaptic CREB phosphorylation during multiple forms of hippocampal synaptic plasticity. *Neuron* **16**, 89–101 (1996).
6. Kavalali, E., Zhuo, M., Bito, H. & Tsien, R. W. Dendritic Ca^{2+} channels characterized by recordings from isolated hippocampal dendritic segments. *Neuron* **18**, 651–663 (1997).
7. Bito, H., Deisseroth, K. & Tsien, R. W. CREB phosphorylation and dephosphorylation: a Ca^{2+} - and stimulus duration-dependent switch for hippocampal gene expression. *Cell* **87**, 1203–1214 (1996).
8. Nakamura, Y., Okuno, S., Sato, F. & Fujisawa, H. An immunohistochemical study of Ca^{2+} /calmodulin-dependent protein kinase IV in the rat central nervous system: light and electron microscopic observations. *Neuroscience* **68**, 181–194 (1995).
9. Sheng, M., Thompson, M. A. & Greenberg, M. E. CREB: a Ca^{2+} -regulated transcription factor phosphorylated by calmodulin-dependent kinases. *Science* **252**, 1427–1430 (1991).
10. Dash, P. K., Karl, K. A., Colicos, M. A., Prywes, R. & Kandel, E. R. cAMP response element-binding protein is activated by Ca^{2+} /calmodulin- as well as cAMP-dependent protein kinase. *Proc. Natl Acad. Sci. USA* **88**, 5061–5065 (1991).
11. Luby-Phelps, K., Hori, M., Phelps, J. M. & Won, D. Ca^{2+} -regulated dynamic compartmentalization of calmodulin in living smooth muscle cells. *J. Biol. Chem.* **270**, 21532–21538 (1995).
12. Pruschy, M., Ju, Y., Spitz, L., Carafoli, E. & Goldfarb, D. S. Facilitated nuclear transport of calmodulin in tissue culture cells. *J. Cell Biol.* **127**, 1527–1536 (1994).
13. Vendrell, M., Pujol, M. J., Tusell, J. M. & Serratos, J. Effect of different convulsants on calmodulin levels and proto-oncogene *c-fos* expression in the central nervous system. *Brain Res. Mol. Brain Res.* **14**, 285–292 (1992).
14. Mitsui, K., Brady, M., Palfrey, H. C. Nairn, A. C. Purification and characterization of calmodulin-dependent protein kinase III from rabbit reticulocytes and rat pancreas. *J. Biol. Chem.* **268**, 13422–13433 (1993).
15. Hahn, K. W., Waggoner, A. S. & Taylor, D. L. A calcium-sensitive fluorescent analog of calmodulin based on a novel calmodulin-binding fluorophore. *J. Biol. Chem.* **265**, 20335–20345 (1990).
16. Westenbroek, R. E., Ahlgrain, M. K. & Catterall, W. A. Clustering of L-type Ca^{2+} channels at the base of major dendrites in hippocampal pyramidal neurons. *Nature* **347**, 281–284 (1990).
17. Magee, J. C. & Johnston, D. Characterization of single voltage-gated Na^{+} and Ca^{2+} channels in apical dendrites of rat CA1 pyramidal neurons. *J. Physiol.* **487**, 67–90 (1995).
18. Morgan, J. I. & Curran, T. Role of ion flux in the control of *c-fos* expression. *Nature* **322**, 552–555 (1986).
19. Murphy, T. H., Worley, P. F. & Barahan, J. M. L-type voltage-sensitive calcium channels mediate synaptic activation of immediate early genes. *Neuron* **7**, 625–635 (1991).

20. Ginty, D. D. et al. Regulation of CREB phosphorylation in the suprachiasmatic nucleus by light and a circadian clock. *Science* **260**, 238–241 (1993).
21. Hagiwara, M. et al. Coupling of hormonal stimulation and transcription via the cyclic AMP responsive factor CREB is rate limited by nuclear entry of protein kinase A. *Mol. Cell. Biol.* **13**, 4852–4859 (1993).
22. Harootyan, A. T. et al. Movement of the free catalytic subunit of cAMP-dependent protein kinase into and out of the nucleus can be explained by diffusion. *Mol. Biol. Cell.* **4**, 993–1002 (1993).
23. Xia, Z., Dudek, H., Miranti, C. K. & Greenberg, M. E. Calcium influx via the NMDA receptor induces immediate early gene transcription by a MAP kinase/ERK-dependent mechanism. *J. Neurosci.* **16**, 5425–5436 (1996).
24. Wang, J., Campos, B., Jamieson, G. A. Jr, Kaetzel, M. A. & Dedman, J. R. Functional elimination of calmodulin within the nucleus by targeted expression of an inhibitor peptide. *J. Biol. Chem.* **270**, 30245–30248 (1995).
25. Hardingham, G. E., Chawla, S., Johnson, C. M. & Bading, H. Distinct functions of nuclear and cytoplasmic calcium in the control of gene expression. *Nature* **385**, 260–265 (1997).
26. Means, A. R. et al. A novel Ca^{2+} /calmodulin-dependent protein kinase and a male germ cell-specific calmodulin-binding protein are derived from the same gene. *Mol. Cell. Biol.* **11**, 3960–3971 (1991).
27. Bito, H., Deisseroth, K. & Tsien, R. W. Ca^{2+} -dependent regulation in neuronal gene expression. *Curr. Opin. Neurobiol.* **7**, 419–429 (1997).
28. Tokumitsu, H. & Soderling, T. R. Requirements for calcium and calmodulin in the calmodulin kinase activation cascade. *J. Biol. Chem.* **271**, 5617–5622 (1996).
29. Sacks, D. B., Porter, S. E., Ladenson, J. H. & McDonald, J. M. Monoclonal antibody to calmodulin: development, characterization, and comparison with polyclonal anti-calmodulin antibodies. *Analyt. Biochem.* **194**, 369–377 (1991).
30. Hulen, D., Baron, A., Salisbury, J. & Clarke, M. Production and specificity of monoclonal antibodies against calmodulin from *Dicotyleum discoideum*. *Cell. Motil. Cytoskel.* **18**, 113–122 (1991).

Acknowledgements. We thank A. Nairn for antibodies; J. Dedman, J. Wang and M. Kaetzel for nCaMBP reagents; D. L. Taylor and J. Montibeller for mCaM; H. Schulman and H. Bito for advice and for comments on the manuscript; and R. Lewis, T. Schwarz and L. Stryer for comments on the manuscript. Supported by grants from the Silvio Conte-NIMH Center for Neuroscience Research, the McKnight Foundation, the Mathers Charitable Trust, and the NIH Medical Scientist Training Program.

Correspondence and requests for materials should be addressed to R.W.T. (e-mail: rwt@leland.stanford.edu).

The hyperthermophile chromosomal protein Sac7d sharply kinks DNA

Howard Robinson*, Yi-Gui Gao*, Bradford S. McCrary†, Stephen P. Edmondson†, John W. Shriver† & Andrew H.-J. Wang*

* Department of Cell and Structural Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA

† Department of Medical Biochemistry, School of Medicine, Southern Illinois University, Carbondale, Illinois 62901, USA

The proteins Sac7d and Sso7d belong to a class of small chromosomal proteins from the hyperthermophilic archaeon *Sulfolobus acidocaldarius* and *S. solfataricus*, respectively^{1,2}. These proteins are extremely stable to heat, acid and chemical agents³. Sac7d binds to DNA without any particular sequence preference and thereby increases its melting temperature by ~40 °C (ref. 4). We have now solved and refined the crystal structure of Sac7d in complex with two DNA sequences to high resolution. The structures are examples of a nonspecific DNA-binding protein bound to DNA, and reveal that Sac7d binds in the minor groove, causing a sharp kinking of the DNA helix that is more marked than that induced by any sequence-specific DNA-binding proteins. The kink results from the intercalation of specific hydrophobic side chains of Sac7d into the DNA structure, but without causing any significant distortion of the protein structure relative to the uncomplexed protein in solution.

Sac7d and Sso7d are proteins of relative molecular mass 7,000 (M_r 7K) from *S. acidocaldarius* and *S. solfataricus*, respectively (Fig. 1a). The structures of Sac7d bound to d(GCGATCGC)₂ and d(GTAATT-TAC)₂ have been well refined at 1.6 and 1.9 Å resolution, respectively (Table 1, Fig. 1b). All ϕ/ψ angles in both complexes fall within the acceptable regions. Sac7d binds to the DNA octamer duplex in the minor groove forming a 1:1 complex (Fig. 1c). In the Sac7d–GCGATCGC complex, the protein binds at the C2pG3 step with a sharp kink, covering four base pairs (G1–C16 to A4–T13) and dramatically widens the DNA minor groove. The other end of the

DNA duplex remains B-DNA-like. In the crystal, one end of the DNA duplex is stacked on the other end of a 2₁-axis symmetry-related duplex, forming an infinite pseudo-DNA duplex along the crystallographic *a*-axis. There are protein–protein contacts between the complexes through the carboxy-terminal α -helices of adjacent proteins. In contrast, the sharp kink in the Sac7d–GTAATTAC complex is located at the A3pA4 step. Despite the different protein-binding sites on DNA, the crystal packing interactions remain unchanged, resulting in isomorphous crystal lattices between the two complexes (Table 1).

The tertiary structure of Sac7d consists of a five-stranded incomplete β -barrel, capped at the opening by a three-turn C-terminal α -helix, the same as the nuclear magnetic resonance (NMR) structure⁵. A notable feature of the β -barrel is the triple-stranded β -sheet (β 3– β 4– β 5), which has been implicated as the DNA-binding surface^{6,7}. The general topology of Sac7d may be considered that of the OB-fold family⁸. There are two 3₁₀ turns that allow the protein's main chain to change direction abruptly. The protein has a small hydrophobic core with only 11 residues. A large number of hydrophobic amino acids are solvent exposed, of which Trp 24, Val 26, Met 29 and Ala 44 are involved in the DNA-binding contacts.

More than half of the surface amino acids are charged (14 lysines, 4 arginines, 7 glutamic acids and 5 aspartic acids). They are not uniformly distributed on the protein surface. The triple-stranded β -sheet that binds DNA has only one positive residue, Arg 42, and no negatively charged amino acids. Many of the negatively charged residues are located on the β 2 strand and the loop between the β 4 and β 5 strand. Such an asymmetric charge may be relevant in the

role of Sac7d in chromatin organization. Only six (Lys 7, 9, 21, 22, 28 and 39) of the 14 lysines are potentially involved in ionic interactions with the phosphates (Fig. 2, top). Of the four arginines, only one (Arg 25) is involved in possible salt-bridge interactions with Glu 47 and Glu 62. The paucity of the observed intramolecular salt bridges argues against the proposal that salt bridges are important for the extreme thermostability of Sac7d.

The binding of Sac7d to DNA involves a binding surface area of about 20 $\times$ 20 Å² (400 Å²) in the minor groove (Fig. 2, bottom). The matching of shapes and charges between Sac7d and DNA is remarkably snug. In both complexes, a buried box-shaped cavity

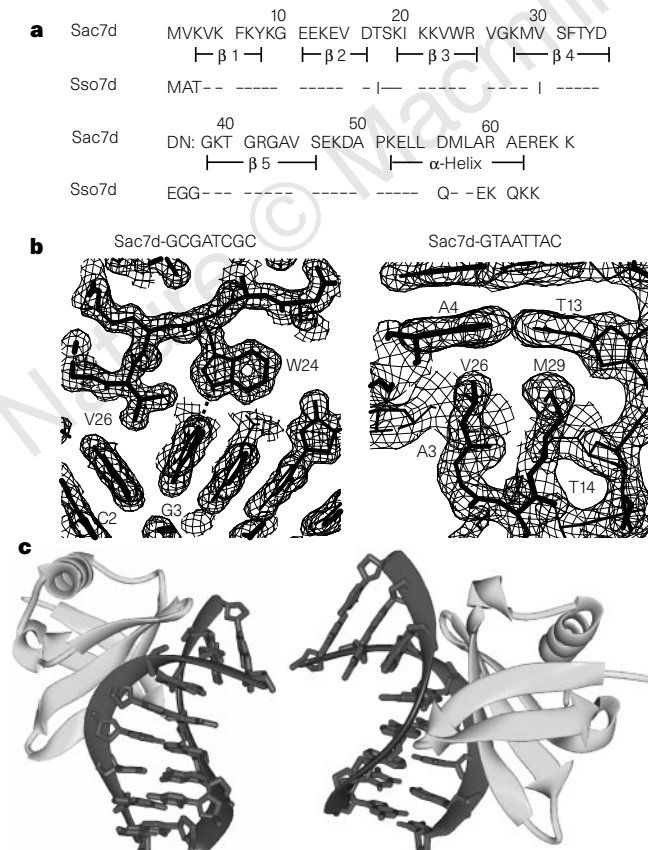


Figure 1 Sequence of Sac7d and Sso7d, and structure of Sac7d–DNA complexes. **a**, Amino-acid sequences of recombinant Sac7d and Sso7d. The first methionine is from the initiation codon. **b**, The ($2F_o - F_c$) Fourier electron density maps (contoured at 1 σ level) of the regions at the interface of protein and DNA. **c**, Ribbon diagrams of the structure of two Sac7d–DNA complexes drawn by MIDAS.²⁷

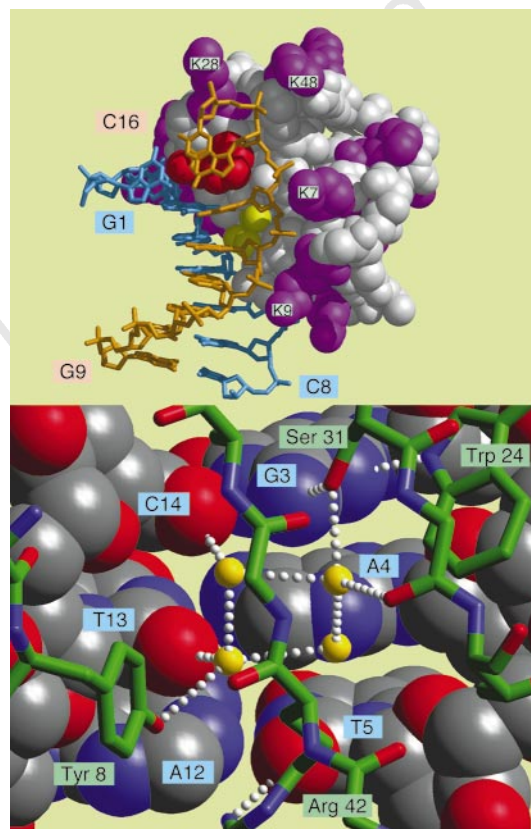


Figure 2 The structure of the Sac7d–GCGATCGC complex. Top, the structure reveals the binding of Sac7d (van der Waals drawing with lysines shown in purple) in the minor groove of DNA (stick drawing). The side chains of the two intercalating amino acids Val26/Met29 and the four bridging waters are shown in red and yellow, respectively. Bottom, the structure surrounding the four bridging waters located in the space between the protein surface and the G3–C14 and A4–T13 base pairs. The single tryptophan (β 3-Trp 24) packs up its bulky ring against the deoxyribose of G3 and A4, filling up the space between DNA and Sac7d, and its indole NH group forms a specific hydrogen bond to the N3 of the G3 base, anchoring the G3 base in its open position. The interactions of this tryptophan explain the observation that it causes the 90% fluorescence quenching when DNA binds to Sac7d⁴. Ser31, which is inaccessible to solvent in the free protein as a result of shielding by Trp24, Val26, Met29 and Ala44, is making a hydrogen bond to the sulphur atom of Met29 in the Sac7d–GTAATTAC complex. In the Sac7d–GCGATCGC complex, Ser31 receives a hydrogen bond (2.87 Å) from the N2 amino group of the G3 nucleotide. The guanidinium group of Arg42 is hydrogen bonded to the O2 atom of T5. The position of Arg42 is held in its place by Tyr8. The hydrogen bond between Arg42 and the O2 atom of a thymine appears to be important, and may determine the polarity and sequence location of the Sac7d binding. The structure of Sac7d complexed to the heteroduplex of GCGAACGC and GCGTTGCG showed that Sac7d is bound to DNA with the Arg42 hydrogen bonded to the T5 of the TT strand, not to the AA strand (results not shown). Attempts to crystallize the complex of Sac7d–GCGTACGC were unsuccessful, suggesting that the thymine at position 5 is important.

Table 1 Crystal and refinement data of two Sac7d–DNA complexes

	Sac7d + GCGATCGC	I-dU-05	I-dC-02	Sac7d + GCGATCGC	Sac7d + GTAATTAC
Crystal data					
<i>a</i> (Å)	51.47	51.45	52.02	50.11	51.77
<i>b</i> (Å)	77.85	78.10	77.08	77.17	77.53
<i>c</i> (Å)	35.97	36.16	36.00	35.48	36.52
Resolution (Å)	2.5	2.5	1.9	1.6	1.9
No. of reflections (>1.0σ(<i>I</i>))	4,848	4,984	9,058	18,497	10,085
Temperature (°C)	20	20	20	–150	20
<i>R</i> _{merge} (%)	4.48	6.59	7.44	7.00	5.74
Completeness (%)	70.6	91.7	76.4	95.0	87.9
Wilson <i>B</i> -factor (Å ²)			34.2	19.4	38.8
Mean overall figure of merit	0.57				
Refinement data					
No. of reflections (>3.0σ(<i>F</i>))				14,584	8,105
<i>R</i> -factor/ <i>R</i> -free (5% data)				0.211/0.252	0.186/0.238
R.m.s.d. bond distance (Å) (Sac7d/DNA)				0.014/0.008	0.013/0.008
R.m.s.d. bond angle (deg) (Sac7d/DNA)				1.58/1.20	1.65/1.28
No. of atoms (Sac7d/DNA)				533/322	533/322
No. of waters				132	68

(6 Å in edge) is situated between the protein and DNA surfaces (for example, near G3–C14 and A4–T13 in Sac7d–GCGATCGC) in which four well ordered water molecules are located. This water-filled cavity is important in that it allows a G–C base pair to be bound without the additional N2 amino group, causing steric problems, thus permitting the protein its sequence-general binding to DNA.

The two most conspicuous features of the nonspecific DNA complex are the kinking of the DNA and the importance of hydrophobic interactions (which may contribute to the change in heat capacity of binding). The latter is consistent with the analysis of the energetics of the complex formation⁴. The DNA is kinked by 61° at the C2pG3 step in the Sac7d–GCGATCGC complex through the intercalation (wedging) of specific amino acids. This type of conformational distortion is analogous to that found in the structure of the complex between the eukaryotic TBP and the TATA-box DNA^{9,10}, and in the DNA complexes of LEF-1 (ref. 11) and SRY (ref. 12), two HMG-box-containing proteins. Remarkably, both Sac7d and TBP use amino acids on a β-sheet for the intercalation, whereas LEF-1 (ref. 11) and SRY (ref. 12) use the amino acids located at the corner of the α-helical L-shape HMG-box. Despite the different protein motifs, the induced DNA deformation is similar in the four structures. The 61° single-step kink in the Sac7d–DNA complexes is the largest known in a protein–DNA complex.

The DNA distortion is associated with helix unwinding (–12° to –14°) at steps surrounding the intercalation sites. There is also a slight roll (11°) between the G3–C14 and A4–T13 base pairs, making a total DNA bend of 72°. The adaptation of the DNA duplex to the irregular surface of Sac7d requires numerous changes in DNA. Many nucleotides surrounding the wedge site adopt the less common C3′-endo (N-type) sugar puckers. Although the conformation of bound DNA resembles A-DNA (as in an RNA duplex), the O2′ hydroxyl groups of the RNA duplex would pose serious van der Waals clashes among the sugars with S-type puckers. Therefore, it seems reasonable to assume that Sac7d would not bind RNA duplex.

The positioning for these intercalating residues is provided by the alignment of the triple-stranded β-sheet with respect to DNA using specific hydrogen bonds or salt bridges to phosphate (Fig. 3). The number of salt bridges between the protein and DNA is in excellent agreement with the five ionic interactions predicted by the salt back-titrations of the complex⁴ using the theory of de Haseth *et al.*¹³.

The structures of the Sac7d–DNA complexes offer insight into the possible role of several DNA-binding proteins, including TBP (refs 9, 10), LEF1 (ref. 11), SRY (ref. 12) and PurR (ref. 14), in transcription regulation. These proteins bind in the minor groove of DNA and kink the DNA duplex to varying degrees. Recognizing the

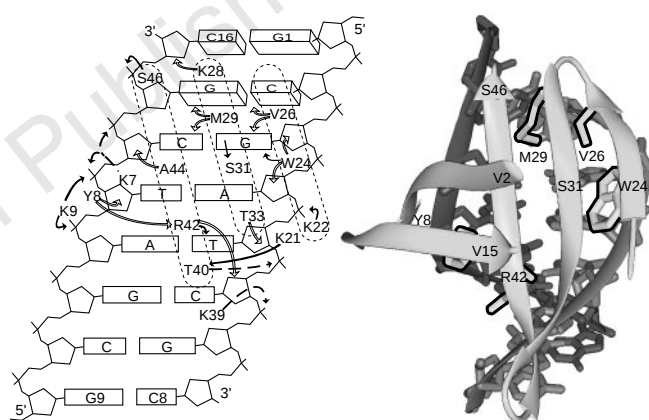


Figure 3 Sac7d–DNA contacts. Left, schematic diagram summarizing all Sac7d–DNA contacts. The filled, open and dashed arrows represent direct hydrogen bonds or salt bridges, respectively. Right, the protein–DNA contacts of the Sac7d–GCGATCGC complex using the triple-stranded β-sheet of Sac7d. Selected side chains of Sac7d are shown.

triple-stranded β-motif as a DNA-binding module in Sac7d and its similarity to the OB-fold⁸ leads us to note that the cold-shock proteins CspB¹⁵ and CspA¹⁶ have similar folding and critical DNA-binding amino acids to those of Sac7d. Both CspA and CspB are related to the Y-box proteins, a widespread and highly conserved structural (sequence) motif occurring from bacterium to human¹⁷. Therefore this structural alignment among those proteins may be significant in understanding the Y-box proteins.

Previous studies have shown that Sac7d binds to double-stranded DNA with a site size of 3.3 to 4.6 base pairs. The structure of the complex presented here is consistent with this and a model of the polymeric Sac7d–DNA complex is shown in Fig. 4.

Our results indicate a second type of chromatin structure in the Archaea. Histone or histone-like proteins (such as HMf)¹⁸ form a core which DNA wraps around¹⁹, but alternatively Sac7d and bacterial chromatin HU protein may bind to DNA in the minor groove and form a higher-ordered structure^{20,21}. (HU may organize the chromatin using the minor-groove binding mode through intercalation, as revealed by the crystal structure of a highly analogous protein, integration host factor (IHF), bound to DNA²².) Thus, there seem to be two different types of DNA-compaction mechanisms in the Archaea: the mechanism we have described using Sac7d, which might be representative of the Crenarchaeota, the organisms closest to the root of the universal



Figure 4 Sac7d binding to DNA duplex. Top, a model of a single Sac7d bound to a long B-DNA duplex by extending the crystal structure (white) at both ends shows the sharp kink induced by Sac7d. Bottom, a model of two turns (8 Sac7d proteins plus 29 base pairs) of the right-handed helix of the polymeric Sac7d-double-stranded DNA complex. The model was built with a maximum binding density of 3 base pairs per Sac7d using the same binding mode as shown above. All Sac7d proteins are oriented with the same polarity. The pitch and the diameter of the helix are 30 and 50 Å, respectively. The coil-like double-stranded DNA (green and gold), resembling a distorted A-DNA, is located close to the centre of the helical complex. The protein shell protects the DNA from the solvent (and therefore from the nuclease attack). Sac7d may have slight sequence preference owing to specific interactions between protein and DNA, for example Arg 42 binding to the O2 of a thymine. This may result in variations of the number of base pairs per Sac7d. In this model the compaction ratio is ~1.2. The polymer complexes may associate with one another using the exterior proteins through Sac7d-Sac7d interactions or through other proteins (such as Sac8 or Sac10). Such association would provide the further compaction of DNA needed for the chromatin structure.

phylogenetic tree²³; and a mechanism involving true histone-like structure²⁰ for the HMf-class of proteins found in the Euryarchaeota. To our knowledge, all of the true histone-like proteins described from the Archaea have been from representatives of the Euryarchaeota. Further work will be needed to determine whether these two chromatin structures are different in the two archaean kingdoms²⁴.

In conclusion, the structural determination of the complexes of Sac7d and DNA oligomers provides examples of the nonspecific DNA-binding mode of a chromosomal protein (Sac7d) and leads to the identification of a DNA-binding structural motif in the Archaea. The remarkable ability of a small but stable protein to sharply kink DNA may have implications in protein design. Further studies of Sac7d and other hyperthermophile DNA-binding proteins (such as Sac8, Sac10 and Sso7d) may reveal how these proteins function in chromatin organization and gene regulation. □

Methods

Collection of diffraction data. The recombinant Sac7d protein was prepared and purified as described¹. The purified protein was dialysed against de-ionized water and lyophilized. The Sac7d-DNA complexes have been crystallized using

the vapour-diffusion method. The Sac7d-GTAATTAC complex was crystallized from 1.3 mM Sac7d, 1.3 mM DNA duplex, 2 mM Tris.Cl buffer, pH 6.5, 2.6% PEG400 solution, equilibrated with 15% PEG400. The Sac7d-GCGATCGC complex and the two iodo derivatives were crystallized from 1.3 mM Sac7d, 1.3 mM DNA duplex, 2 mM Tris.Cl buffer, pH 6.5, 5% MPD solution, equilibrated with 20% MPD. Diffraction data were collected either at ambient temperature or at -150 °C on a Rigaku R-Axis Ilc Image-plate area detector system. The crystals of both complexes are in the space group $P2_12_12_1$ with the unit cell dimensions shown in Table 1. The data have been processed using software from Molecular Structure Corporation.

Phase determination. The phases were determined by the multiple isomorphous replacement (MIR) method using two iodo derivatives (denoted as I-dC-02 and I-dU-05 with the iodine located at positions C2 and T5, respectively) for the Sac7d-GCGATCGC complex listed in Table 1. The figure-of-merit weighted MIR map at 2.5 Å resolution clearly revealed both the DNA and the Sac7d protein electron density. The structures of both complexes were refined using data beyond 6 Å resolution by the simulated annealing procedure incorporated in X-PLOR²⁵. Improved DNA force-field parameters²⁶ were used. Well ordered water molecules were located using the difference Fourier maps.

Received 29 July; accepted 24 November 1997.

- McAfee, J. G. *et al.* Gene cloning, expression, and characterization of the Sac7 proteins from the hyperthermophile *Sulfolobus acidocaldarius*. *Biochemistry* **34**, 10063-10077 (1995).
- Choi, T., Wittmann-Liebold, B. & Reinhardt, R. Microsequence analysis of DNA-binding proteins 7a, 7b and 7c from the archaeobacterium *Sulfolobus acidocaldarius*. *J. Biol. Chem.* **263**, 7087-7093 (1988).
- McCrary, B. S., Edmondson, S. P. & Shriver, J. W. Hyperthermophile protein folding thermodynamics: differential scanning calorimetry and chemical denaturation of Sac7d. *J. Mol. Biol.* **264**, 784-805 (1996).
- McAfee, J. G., Edmondson, S. P., Zegar, I. & Shriver, J. W. Equilibrium DNA binding of Sac7d protein from the hyperthermophile *Sulfolobus acidocaldarius*: fluorescence and circular dichroism studies. *Biochemistry* **35**, 4034-4045 (1996).
- Edmondson, S. P., Qiu, L. & Shriver, J. W. Solution structure of the DNA-binding protein Sac7d from the hyperthermophile *Sulfolobus acidocaldarius*. *Biochemistry* **34**, 13289-13304 (1995).
- Baumann, H. *et al.* Solution structure and DNA-binding properties of a thermostable protein from the archaeon *Sulfolobus solfataricus*. *Nature Struct. Biol.* **1**, 808-819 (1994).
- Baumann, H. *et al.* DNA-binding surface of the Sso7d protein from *Sulfolobus solfataricus*. *J. Mol. Biol.* **247**, 840-846 (1995).
- Murzin, A. G. OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences. *EMBO J.* **12**, 861-867 (1993).
- Kim, Y., Geiger, J. H., Hahn, S. & Sigler, P. B. Crystal structure of a yeast TBP/TATA-box complex. *Nature* **365**, 512-520 (1993).
- Kim, J. L., Nikolov, D. B. & Burley, S. K. Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature* **365**, 520-527 (1993).
- Love, J. J. *et al.* Structural basis for DNA bending by the architectural transcription factor LEF-1. *Nature* **376**, 791-795 (1995).
- Werner, M. H., Huth, J. R., Gronenborn, A. M. & Clore, G. M. Molecular basis of human 46X, Y sex reversal revealed from the three-dimensional solution structure of the human SRY-DNA complex. *Cell* **81**, 705-714 (1995).
- de Haseth, P., Lohman, T. & Record, M. T. Jr Nonspecific interaction of the lac repressor with DNA: an association reaction driven by counterion release. *Biochemistry* **16**, 4783-4790 (1977).
- Schumacher, M. A., Choi, K. Y., Zalkin, H. & Brennan, R. G. Crystal structure of LacI member, PurR, bound to DNA: minor groove binding by alpha helices. *Science* **266**, 763-770 (1994).
- Schindelin, H., Marahiel, M. A. & Heinemann, U. Universal nucleic acid-binding domain revealed by crystal structure of the *B. subtilis* major cold-shock protein. *Nature* **364**, 164-168 (1993).
- Schindelin, H., Jiang, W., Inouye, M. & Heinemann, U. Crystal structure of CspA, the major cold shock protein of *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **91**, 5119-5123 (1994).
- Wolfe, A. P. Structural and functional properties of the evolutionarily ancient Y-box family of nucleic acid binding proteins. *BioEssays* **16**, 245-251 (1994).
- Starich, M. R., Sandman, K., Reeve, J. N. & Summers, M. F. NMR structure of HMfB from the hyperthermophile, *Methanothermobacter fervidus*, confirms that this archaeal protein is a histone. *J. Mol. Biol.* **255**, 187-203 (1996).
- Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251-260 (1997).
- Grayling, R. A., Sandman, K. & Reeve, J. Histones and chromatin structure in hyperthermophilic Archaea. *FEMS Microbiol. Rev.* **18**, 203-213 (1996).
- Dijk, J. & Reinhardt, R. in *Bacterial Chromatin* (eds Gualerzi, C. O. & Pon, C. L.) Springer-Verlag, Berlin, pp. 185-218 (1986).
- Rice, P. A., Yang, S., Mizuuchi, K. & Nash, H. A. Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. *Cell* **87**, 1295-1306 (1996).
- Woese, C. R., Kandler, O. & Wheelis, M. L. Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl Acad. Sci. USA* **87**, 4576-4579 (1990).
- Ronimus, R. & Musgrave, D. R. A comparison of the DNA binding properties of histone-like proteins derived from representatives of two kingdoms of the archaea. *FEMS Microbiol. Lett.* **134**, 79-84 (1995).
- Brünger, A. T. *X-PLOR 3.1, A System for X-ray Crystallography and NMR* (Yale Univ. Press, New Haven, CT, 1992).
- Parkinson, G., Vojtechovsky, J., Clowney, L., Brunger, A. T. & Berman, H. M. New parameters for the refinement of nucleic acid-containing structures. *Acta Crystallogr.* **52**, 57-64 (1996).
- Ferrin, T. E., Huang, C. C., Jarvis, L. E. & Langridge, R. The MIDAS display system. *J. Mol. Graphics* **6**, 13-27 (1988).

Acknowledgements. This work was supported by the NIH.

Correspondence and requests for materials should be addressed to A.H.-J.W. (e-mail: ahjwang@uiuc.edu). The atomic coordinates have been deposited in the Brookhaven Protein Database (accession nos. 1AZP and 1AZQ).

Structure at 0.85 Å resolution of an early protein photocycle intermediate

Ulrich K. Genick, S. Michael Soltis*, Peter Kuhn*, Ilona L. Canestrelli & Elizabeth D. Getzoff

Department of Molecular Biology and Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA

* Stanford Synchrotron Radiation Laboratory, PO Box 4349, Bin 69, Stanford, California 94309, USA

Protein photosensors from all kingdoms of life^{1,2} use bound organic molecules, known as chromophores, to detect light. A specific double bond within each chromophore is isomerized by light, triggering slower changes in the protein as a whole. The initial movements of the chromophore, which can occur in femtoseconds, are tightly constrained by the surrounding protein, making it difficult to see how isomerization can occur, be recognized, and be appropriately converted into a protein-wide structural change and biological signal. Here we report how this

dilemma is resolved in the photoactive yellow protein (PYP). We trapped a key early intermediate in the light cycle of PYP at temperatures below -100°C , and determined its structure at better than 1 Å resolution. The 4-hydroxycinnamoyl chromophore^{3,4} isomerizes by flipping its thioester linkage with the protein, thus avoiding collisions resulting from large-scale movement of its aromatic ring during the initial light reaction. A protein-to-chromophore hydrogen bond that is present in both the preceding dark state⁵ and the subsequent signalling state⁶ of the photosensor breaks, forcing one of the hydrogen-bonding partners into a hydrophobic pocket. The isomerized bond is distorted into a conformation resembling that in the transition state. The resultant stored energy is used to drive the PYP light cycle. These results suggest a model for phototransduction, with implications for bacteriorhodopsin^{7,8}, photoactive proteins^{1,2}, PAS domains⁹, and signalling proteins.

PYP¹⁰ is a cytosolic light receptor that is thought to mediate the negative phototactic response to intense blue light¹¹ in certain halophilic bacteria. PYP's buried chromophore, a *p*-hydroxycinnamoyl anion covalently bound to the side chain of Cys 69 via a thioester linkage^{3,4,12}, gives PYP's ground state its bright yellow colour (λ_{max} 446 nm). After the initial light-driven reaction, PYP undergoes a series of spontaneous spectroscopic changes^{13–15}, leading to the formation of the colourless (λ_{max} 355 nm), longest lived intermediate I2, which is presumed to be the signalling state of PYP⁶. From

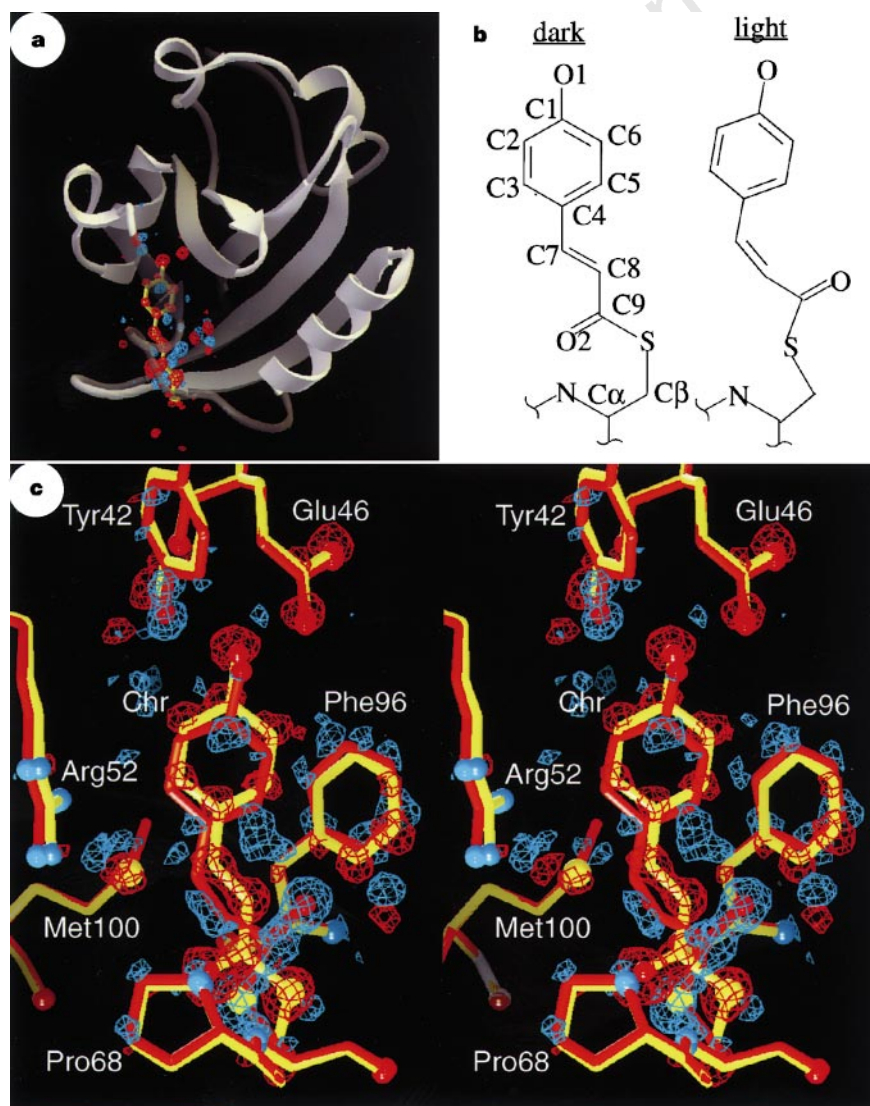


Figure 1 Structural consequences of light activation.

a, Overview showing that electron-density differences (decrease in red basket weave, increase in blue basket weave) are restricted to the immediate vicinity of the chromophore (yellow stick figure). The overall fold is shown as a white ribbon. The $F_{\text{light}} - F_{\text{dark}}$ difference map phased from the 1.4-Å ground-state model⁵ is contoured at 5σ . **b**, Chemical drawings of the chromophore in the ground (dark; left) and light-activated (right) states. **c**, Stereo view of the active site showing refined models for the dark state (yellow) and the early intermediate (orange), with the difference electron density map described in **a** contoured at 4σ . The difference map shows that chromophore isomerization took place by flipping the thioester linkage (lower centre) between the chromophore (Chr) and the protein. The difference map also shows minor adjustments in the surrounding protein.

there, PYP returns to the ground (initial) state in a protein-mediated thermal process¹⁶, completing the light cycle. Between the structures of the ground, or dark, state⁵ and the bleached signalling state, the central double bond of the chromophore undergoes a *trans*-to-*cis* isomerization and the phenolic oxygen (O1; see Fig. 1 for nomenclature) moves from the hydrophobic core of the protein to become solvent-exposed and protonated⁶. The resulting changes in the shape and the electrostatic and hydrogen-bonding properties of the protein's surface are thought to constitute the signal to PYP's downstream partner⁶.

The high speed of the initial photoisomerization reaction, the tight packing of hydrophobic cores in proteins, and the ultimate need to undergo a significant structural change to generate a biological signal place seemingly contradictory demands on protein photoreceptors. To understand how the protein accommodates rapid isomerization of the chromophore without collisions, and then amplifies this restricted movement into the larger structural changes required for signalling, we determined the X-ray crystallographic structure of an early photocycle intermediate. We cryogenically cooled and then light-activated a PYP crystal. After light activation, PYP progresses through its photocycle by using thermal energy to cross energy barriers. Lowering the temperature limits the height of energy barriers that can be crossed, causing the photocycle

to arrest when it first encounters a newly insurmountable energy barrier. Cryotrapping of PYP solutions at different temperatures allows the accumulation and spectroscopic characterization of new intermediates of the PYP light cycle¹⁵ that precede early intermediate I1, and thus exceed the nanosecond time-resolution of the room-temperature spectroscopic studies performed to date^{13,14}.

We collected two complete diffraction data sets (Table 1) at very high resolution from a single PYP crystal kept continuously cooled in a stream of evaporating nitrogen. The first data set (F_{dark}) was collected in the absence of light on a dark-adapted crystal, and the second (F_{light}) was collected following activation with light of 460 nm wavelength (see Methods). Difference electron density maps of $F_{\text{dark}} - F_{\text{light}}$ (Fig. 1) showed that structural changes were restricted to the chromophore and its immediate environment, and had not yet progressed to the protein's surface. This confirmed that an early intermediate of the chromophore had accumulated. Electron-density maps also showed how the chromophore avoids collisions during the initial, rapid, light-driven isomerization of the chromophore. By flipping the thioester bond with Cys 69, instead of moving the aromatic ring, the number of atoms that move and the distance they travel are minimized (Fig. 1b). The temperature and illumination conditions used here and preliminary spectroscopic studies of PYP crystals (see Methods) indicate that the observed light-activated structure corresponds to the early intermediate termed PYP_{BL} by Imamoto *et al.*¹⁵.

Diffraction data at <1 Å resolution (Table 1) collected at beam-line 9-1 of the Stanford Synchrotron Radiation Laboratory (F_{dark} 0.82 Å and F_{light} 0.85 Å resolution) allowed full matrix crystallographic refinement in SHELX-97 (ref. 17) of atomic positions and anisotropic thermal parameters with minimal reliance on stereochemical restraints, as well as determination of estimated standard deviations for all parameters. The high quality and completeness of F_{light} (Table 1) allowed unambiguous structure determination of the light-activated intermediate (Fig. 1c), despite the expected low occupancy of this intermediate (refined to 25%) and its extensive structural overlap with the ground-state structure which is still occupied by most (refined to 75%) of the molecules in the light-activated crystal. A model for the isomerized chromophore was built into the difference electron density map (Fig. 1c) and then added to the dark-state model (previously refined against F_{dark}) as an alternative conformer with partial occupancy. The combined models were then refined against F_{light} . After several cycles of model building and refinement, 11 amino-acid residues (Tyr 42, Glu 46, Arg 52, Val 66, Ala 67, Pro 68, Cys 69, Thr 70, Phe 96, Asp 97, Met 100) and one water molecule (H₂O 1,004) shifted during

Table 1 Diffraction data and refinement statistics

	Dark		Light	
Data set				
Resolution (Å)	0.82		0.85	
Observations	613,864		561,693	
Unique reflections	94,197		84,934	
Completeness (%)				
Overall	97.5		97.7	
Highest shell	85.7 (0.83–0.82 Å)		87.1 (0.86–0.85 Å)	
R_{sym}^* (%)	3.1		4.2	
Refinement				
	$F > 4\sigma$	All	$F > 4\sigma$	All
$R_{\text{free}}^\dagger$ (%)	13.36	14.35	14.28	15.48
$R_{\text{work}}^\dagger$ (%)	11.64	12.46	12.13	13.26
R.m.s.d.‡				
Bond length (Å)	0.020		0.017	
Angular distances (Å)	0.043		0.038	

* $R_{\text{sym}} = \sum_i \sum_j |I_{ij} - \langle I_{ij} \rangle| / \sum_i \sum_j I_{ij}$.

† Calculated with 5% of random reflections excluded from refinement.

‡ R.m.s.d., root-mean-square deviation.

§ SHELX-97 defines angles in distances between an angle's flanking atoms.

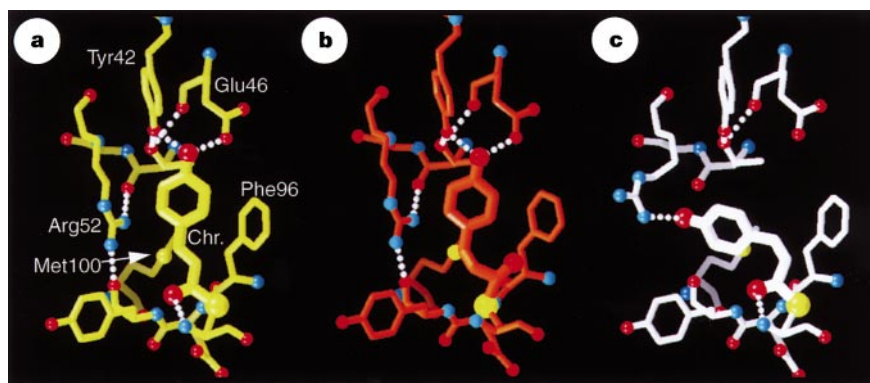


Figure 2 Active-site structures of the (a) ground state, (b) early intermediate, and (c) bleached state of PYP. During the very rapid transition from the ground state (a) to the early intermediate (b), movements in the light-driven *trans*-to-*cis* isomerization of the *p*-hydroxycinnamoyl chromophore (Chr., thick bonds, centre) are restricted to the flipping of the thioester linkage. The hydrogen-bonding network (dotted lines) to the phenolic oxygen at the tip of the

chromophore (top) is preserved, whereas the hydrogen bond between the chromophore's thioester oxygen and the backbone amide is disrupted (bottom). During the slower transition to the bleached signalling state (c), the chromophore retains the *cis* configuration while re-forming this hydrogen bond. Arg 52 (left) is displaced, exposing the chromophore's tip to solvent and generating a structural signal on the protein's surface⁶.

refinement and showed paired positive and negative peaks in difference electron density maps (Fig. 1c). These residues were added to the model of the light-activated state and refined to convergence.

The resulting structures for the light-activated intermediate and the ground state allow detailed analysis of the isomerized chromophore conformation and its influence on the surrounding protein environment (Figs 2 and 3). Interestingly, the cryotrapped, light-activated structure shows the flipping of the thioester linkage to be virtually complete (166° rotation of the thioester carbonyl occurred relative to the plane of the aromatic ring), whereas the isomerizable double bond of the chromophore has barely crossed (torsion angle of -80°) the *trans*-to-*cis* transition point. The chromophore thus has a distorted transition-state-like conformation (Table 2). Thus the flipping of the thioester linkage, which involves the largest atomic movements in the early light reaction ($3.4 \text{ \AA}$ for O2 and $1.4 \text{ \AA}$ for S), precedes the full isomerization of the double bond. We propose that the flipping of the thioester linkage forms the committed step that guides the double bond's relaxation from the transition state towards the *cis* conformation. This protein-mediated guidance could explain why the isomerizable bond, once it reaches the transition state, does not return to the *trans* conformation that perfectly matches the protein environment but instead forms the *cis* conformation, causing the mismatch with the protein environment that drives the rest of the light cycle.

The high-resolution structures show how the protein accommodates the flipped thioester linkage with a few, very small adjust-

ments, while conserving most of the active-site hydrogen bonds. The largest atomic shift ($0.5 \text{ \AA}$) elsewhere in the protein occurs in the ring of Phe 96, which swivels slightly to adjust to the new position of the chromophore's thioester oxygen (O2). Two small packing gaps that flank the ring of Phe 96 in the ground-state structure allow this swivel without collisions. Tyr 42 (largest atomic shift $0.31 \text{ \AA}$) and Glu 46 ($0.13 \text{ \AA}$ shift) follow the slight lengthwise contraction (of $0.4 \text{ \AA}$) of the chromophore upon isomerization. These movements preserve the hydrogen bonds (Fig. 2) between these residues and the chromophore's phenolic oxygen (O1), and explain the spectroscopically observed preservation of these hydrogen bonds in intermediate I1 (refs 16, 18). Another consequence of thioester flipping in the early intermediate is disruption of the tight hydrogen bond ($2.6 \text{ \AA}$ length) between the thioester oxygen and the backbone amide of Cys 69. This hydrogen bond is present in the structures of both the ground state⁵ and the bleached intermediate I2 (ref. 6) (Fig. 2). In the structure of the early photocycle intermediate, this hydrogen bond is eliminated and the thioester carbonyl oxygen atom is placed in a completely hydrophobic pocket formed by the aromatic rings of the chromophore and Phe 75, Tyr 94, and Phe 96 (Fig. 3).

Analysis of the crystallographic structures of the ground state, the early intermediate, and the bleached signalling state (Fig. 2), with spectroscopic data, allows us to propose an energy budget for the PYP photocycle. Only the initial step of the PYP photocycle is light-driven, so all subsequent steps must be driven by energy from the absorbed photon (62 kcal of energy is absorbed at 460 nm). In the light-activated early intermediate structure reported here, the isomerizable double bond lies near the transition state, indicating that much of the initial photon energy may still be stored in the highly distorted geometry of the chromophore. Additionally, a presumably smaller amount of energy is stored in the mismatch between the protein conformation and the chromophore configuration. In $<2 \text{ ns}$, the photocycle progresses to intermediate I1 in which roughly half of the photon energy has been released as heat, as determined by laser-induced optoacoustic spectroscopy¹⁹. But infrared spectroscopy¹⁸ and time-resolved optical spectroscopy on site-directed mutants of PYP¹⁶ indicate that intermediate I1 retains many features of the earlier cryotrapped structure reported here (that is, the disruption of the O2-to-backbone hydrogen bond, retention of the O1-to-Glu 46 hydrogen bond and isomerization of the chromophore).

We therefore conclude that the formation of intermediate I1 mainly involves the continuation of the rotation of the double bond, releasing the extreme distortions in the chromophore and forming a fully *cis* conformation. At a timescale of microseconds, protein and chromophore next rearrange around one another to form intermediate I2. These rearrangements re-form the O2-to-backbone hydrogen bond and create a structural signal on the protein surface. In the last step of the PYP light cycle, the chromophore re-isomerizes, releasing the energy stored in its non-planar *cis* conformation⁶. This again triggers protein rearrangements which return PYP to the ground state, at a slow millisecond timescale. This completes the light cycle.

By first cryogenically cooling a protein crystal and then activating it, we cryotrapped an early photocycle intermediate which, at room temperature, has a lifetime of less than 1 ns . Other crystallographic cryotrapping approaches either apply an activate-then-cool strategy^{20,21} (currently limited to a millisecond-to-second timescale by the speed of sample cooling), or use low temperatures to slow down reactions so that continuous activation yields a significant steady-state population of the slowest reaction intermediate²². In the cool-then-activate strategy, control of the crystal temperature determines the point to which the reaction can progress, and thereby allows separation of initial, very fast small changes (for example, in a ligand) from slower, large conformational shifts in proteins. This extends the applicability of crystallographic

Table 2 Chromophore geometry*

	Dark	Light
Angles (e.s.d.†)		
C4-C7-C8	125.14 (1.88)	127.85 (5.07)
C7-C8-C9	123.26 (1.72)	131.20 (5.91)
C8-C9-O2	119.90 (1.67)	112.13 (4.74)
C8-C9-SG	115.48 (1.23)	93.92 (3.22)
O2-C9-SG	124.50 (1.06)	148.62 (3.60)
C9-SG-CB	97.51 (0.83)	116.64 (3.55)
Dihedrals (e.s.d.†)		
C5-C4-C7-C8	0.30 (2.13)	42.85 (8.05)
C4-C7-C8-C9	161.99 (1.41)	-79.64 (7.96)
C7-C8-C9-O2	-11.24 (2.24)	42.57 (6.64)
O2-C9-SG-CB	2.66 (1.45)	88.29 (6.51)
C9-SG-CB-CA	-81.51 (1.75)	51.87 (8.38)

* All values given in degrees.

† Estimated standard deviations (e.s.d.) calculated by SHELX¹⁷.

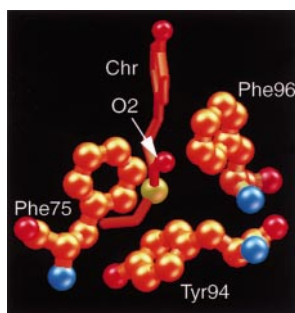


Figure 3 The environment of the thioester oxygen atom (O2) of the chromophore in the early intermediate. Instead of forming a hydrogen bond to the backbone amide of Cys 69 as in the preceding dark state (Fig. 2a) and subsequent signalling state (Fig. 2c), O2 is located in a hydrophobic pocket formed by three aromatic side chains (orange spheres) and the aromatic ring of the chromophore itself (orange sticks, top). We propose that the resolution of this mismatch and the re-formation of the O2-to-backbone hydrogen bond help to drive the formation of the bleached intermediate I2 (Fig. 2c) later in the photocycle.

cryotrapping to new biological problems, such as allosteric control of proteins, light sensing and signalling, which were previously accessible only by the technically more challenging, time-resolved Laue method^{23,24}. Here, cryocooling-then-activating reveals in atomic detail how PYP has achieved minimal atomic movement during the initial events in protein photocycles²⁵. By flipping its thioester linkage rather than its aromatic ring, the chromophore isomerizes with minimal perturbation of the tightly packed protein interior. The rapid photochemical reaction is translated first into a higher energy geometry, which can then be amplified by slower protein shifts to create a biological signal with a millisecond lifetime. These cryotrapped high-resolution PYP structures allow detailed theoretical and computational studies and provide the prototype for comparison to other protein photocycles^{1,2,7,8,25} and signalling processes. □

Methods

Protein purification and crystallization. PYP was purified³ and crystallized⁵ as described.

Diffraction data collection. A $200 \times 200 \times 800 \mu\text{m}^3$ PYP crystal was transferred to a 70% saturated solution of $(\text{NH}_4)_2\text{SO}_4$ with 10% v/v glycerol and immediately frozen in a stream of evaporating nitrogen. The temperature, measured with a thermocouple, at the location of the crystal was -124°C . Data were collected at SSRL beamline 9-1, which receives 3-mrad radiation from a 16-pole wiggler. A Si₁₁₁ single crystal monochromator was used to select X-rays of 0.77-Å wavelength. Both data sets were recorded on a Mar Research image plate in three passes each, using crystal-to-detector distances of 100, 250 and 300 mm. For the high-resolution pass, the centre of the image plate was protected from overexposure by an oversized beam stop made of 1-mm thick aluminum. From ~10 min before freezing until the end of collection of the F_{dark} data set, the crystal was kept in the dark.

Light activation. After collection of the F_{dark} data set, the crystal was continuously rotated and illuminated with low-level blue light (20-nm band path centred at 460 nm) for 1 h. During collection of the F_{light} data set on the same crystal, the illumination was continued to eliminate possible effects of secondary photoequilibration of the accumulated intermediate¹⁵ by residual ambient light. Light was provided by a 75-W Xenon arc lamp illuminating a 1/8-m monochromator equipped with a 300 L per mm grating and 1.56-mm slits. The monochromator delivered light through a 400 μm quartz fibre imaged on the crystal at a roughly twofold magnification with two PCX quartz lenses. (All components from Oriol Corp, Stratford, CT, USA.)

Microspectrophotometry. The optical density of PYP crystals calculated from the unit cell parameters (space group $P6_3$; one molecule per asymmetric unit; $a = b = 65.70 \text{ \AA}$; $c = 40.24 \text{ \AA}$) and the molar extinction coefficient¹³ ($45.5 \text{ mM}^{-1} \text{ cm}^{-1}$) is $0.25 \mu\text{m}^{-1}$. This requires PYP crystals to be smaller than ~7 μm to give optical densities that can be measured reliably. We built a microspectrophotometer modified from an existing design²⁶ to achieve the small focal spots (<3 μm FWHM (full width at half maximum)) required to record spectra on such tiny crystals. Because of this small crystal size, even minimal drift in the crystal position caused instability in the baseline of the optical spectra, making it impossible to record reliable difference spectra between the dark and the light-activated states. However, we were able to record absorption spectra from light-activated PYP crystals kept at temperatures below -110°C that show the formation of a 400-nm shoulder on the PYP ground-state spectrum; this shoulder is consistent with the accumulation of intermediate PYP_{BL}¹⁵.

Data processing and structural refinement. Data were processed with Denzo and Scalepack²⁷. Refinement was carried out using SHELX-97 (ref. 17). PYP's ground-state structure was refined against F_{dark} starting from the published structure at 1.4 Å resolution⁵. After refinement of anisotropic thermal parameters and the inclusion of ground-state dual conformers for several side chains, backbone hydrogen atoms became visible in $F_{\text{o}} - F_{\text{c}}$ maps. Inclusion of explicit hydrogen atoms and further refinement dropped the R-factor to 11.76%. A model for the isomerized chromophore was built iteratively into $F_{\text{o/light}} - F_{\text{c/dark}}$ and $2F_{\text{o/light}} - F_{\text{c/light}}$ electron-density maps and included as an alternate conformer. After several rounds of refinement, the structure of PYP's protein component refined against F_{light} was

compared with the model refined against F_{dark} . Residues Tyr 42, Glu 46, Arg 52, Val 66, Ala 67, Pro 68, Cys 69, Thr 70, Phe 96, Asp 97 and Met 100 and water molecule 1,004 showed small shifts and showed difference peaks in a $F_{\text{o/dark}} - F_{\text{o/light}}$ map contoured at 4σ . These residues were also modelled with dual conformers corresponding to the dark and light-activated states and the combined model was refined against F_{light} . After convergence of refinement, block-matrix inversion was used to calculate estimated standard deviations for atomic positions¹⁷.

Received 3 October 1997; accepted 5 February 1998.

1. Unger, V. M., Hargrave, P. A., Baldwin, J. M. & Schertler, G. F. X. Arrangement of rhodopsin transmembrane α -helices. *Nature* **389**, 203–206 (1997).
2. Quail, P. H. *et al.* Phytochromes: photosensory perception and signal transduction. *Science* **268**, 675–680 (1995).
3. Baca, M. *et al.* Complete chemical structure of photoactive yellow protein: novel thioester-linked 4-hydroxycinnamyl chromophore and photocycle chemistry. *Biochemistry* **33**, 14369–14377 (1994).
4. Hoff, W. D. *et al.* Thiol ester-linked p-coumaric acid as a new photoactive prosthetic group in a protein with rhodopsin-like photochemistry. *Biochemistry* **33**, 13959–13962 (1994).
5. Borgstahl, G. E. O., Williams, D. R. & Getzoff, E. D. 1.4 Å structure of photoactive yellow protein, a cytosolic photoreceptor: unusual fold, active site, and chromophore. *Biochemistry* **34**, 6278–6287 (1995).
6. Genick, U. K. *et al.* Structure of a protein photocycle intermediate by millisecond time-resolved crystallography. *Science* **275**, 1471–1475 (1997).
7. Pebay-Peyroula, E., Rummel, G., Rosenbusch, J. P. & Landau, E. M. X-ray structure of bacteriorhodopsin at 2.5 Å from microcrystals grown in lipidic cubic phases. *Science* **277**, 1676–1681 (1997).
8. Kimura, Y. *et al.* Surface of bacteriorhodopsin revealed by high-resolution electron crystallography. *Nature* **389**, 206–211 (1997).
9. Kay, S. A. PAS, present, and future: clues to the origins of circadian clocks. *Science* **276**, 753–754 (1997).
10. Meyer, T. E. Isolation and characterization of soluble cytochromes, ferredoxins and other chromophoric proteins from the halophilic phototrophic bacterium *Ectothiorhodospira halophila*. *Biochim. Biophys. Acta* **806**, 175–183 (1985).
11. Sprenger, W. W., Hoff, W. D., Armitage, J. P. & Hellingwerf, K. J. The eubacterium *Ectothiorhodospira halophila* is negatively phototactic, with a wavelength dependence that fits the absorption spectrum of the photoactive yellow protein. *J. Bacteriol.* **175**, 3096–3104 (1993).
12. Kim, M., Mathies, R. A., Hoff, W. D. & Hellingwerf, K. J. Resonance Raman evidence that the thioester-linked 4-hydroxycinnamyl chromophore of photoactive yellow protein is deprotonated. *Biochemistry* **34**, 12669–12672 (1995).
13. Meyer, T. E., Tollin, G., Hazzard, J. H. & Cusanovich, M. A. Photoactive yellow protein from the purple phototrophic bacterium, *Ectothiorhodospira halophila*. Quantum yield of photobleaching and effects of temperature, alcohols, glycerol, and sucrose on kinetics of photobleaching and recovery. *Biophys. J.* **56**, 559–564 (1989).
14. Hoff, W. D. *et al.* Measurement and global analysis of the absorbance changes in the photocycle of the photoactive yellow protein from *Ectothiorhodospira halophila*. *Biophys. J.* **67**, 1691–1705 (1994).
15. Imamoto, Y., Kataoka, M. & Tokunaga, F. Photoreaction cycle of photoactive yellow protein from *Ectothiorhodospira halophila* studied by low-temperature spectroscopy. *Biochemistry* **35**, 14047–14053 (1996).
16. Genick, U. K. *et al.* Active site mutants implicate key residues for control of color and light cycle kinetics of photoactive yellow protein. *Biochemistry* **36**, 8–14 (1997).
17. Sheldrick, G. M. in *Crystallographic Computing 7* (eds Watenpaugh, K. & Bourne, P.) (Oxford Press, Oxford, in the press).
18. Xie, A., Hoff, W. D., Kroon, A. R. & Hellingwerf, K. J. Glu46 donates a proton to the 4-hydroxycinnamate anion chromophore during the photocycle of photoactive yellow protein. *Biochemistry* **35**, 14671–14678 (1996).
19. van Brederode, M. E., Gensch, T., Hoff, W. D., Hellingwerf, K. J. & Braslavsky, S. E. Photoinduced volume change and energy storage associated with the early transformations of the photoactive yellow protein from *Ectothiorhodospira halophila*. *Biophys. J.* **68**, 1101–1109 (1995).
20. Stowell, M. H. B. *et al.* Light-induced structural changes in photosynthetic reaction center: implications for mechanism of electron-proton transfer. *Science* **276**, 812–816 (1997).
21. Williams, P. A. *et al.* Haem-ligand switching during catalysis in crystals of a nitrogen-cycle enzyme. *Nature* **389**, 406–412 (1997).
22. Schlichting, I., Berendzen, J., Phillips, G. N. Jr & Sweet, R. M. Crystal structure of photolysed carbonmonoxy-myoglobin. *Nature* **371**, 808–812 (1994).
23. Hajdu, J. & Anderson, I. Fast crystallography and time-resolved structures. *Annu. Rev. Biophys. Biomol. Struct.* **22**, 467–498 (1993).
24. Bourgeois, D. *et al.* Feasibility and realization of single-pulse Laue diffraction on macromolecular crystals at ESRF. *J. Synchrotron Rad.* **3**, 65–74 (1996).
25. Schertler, G. F. X., Lozier, R., Michel, H. & Oesterheld, D. Chromophore motion during the bacteriorhodopsin photocycle: polarized absorption spectroscopy of bacteriorhodopsin and its M-state in bacteriorhodopsin crystals. *EMBO J.* **10**, 2353–2361 (1991).
26. Hadfield, A. & Hajdu, J. A fast and portable microspectrophotometer for protein crystallography. *J. Appl. Cryst.* **26**, 839–842 (1993).
27. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).

Acknowledgements. We thank C. D. Mol and C. D. Putnam for testing the PYP crystals for sub-Ångstrom diffraction; B. E. Honig for discussing low-temperature approaches to the study of photosensors; J. Berendzen for discussing appropriate illumination conditions; C. L. Brooks III, B. N. Dominj, J.-L. Pellequer and D. A. Case for exploring possibilities of energy calculation; and T. Woo for growing the crystal used for microspectrophotometry. This research was supported by a grant from the N.I.H. (to E.D.G.) and a scholarship from Boehringer Ingelheim Fonds (for U.K.G.). This work is based on research conducted at the Stanford Synchrotron Radiation Laboratory (SSRL), which is funded by the Department of Energy, Office of Basic Energy Sciences. The Biotechnology Program is supported by the NIH, National Center for Research Resources, Biomedical Technology Program and the Department of Energy, Office of Biological and Environmental Research.

Correspondence and requests for materials should be addressed to E.D.G. (e-mail: edg@scripps.edu). Crystallographic coordinates were deposited with the Brookhaven Protein Data Bank (accession number BNL-20580).

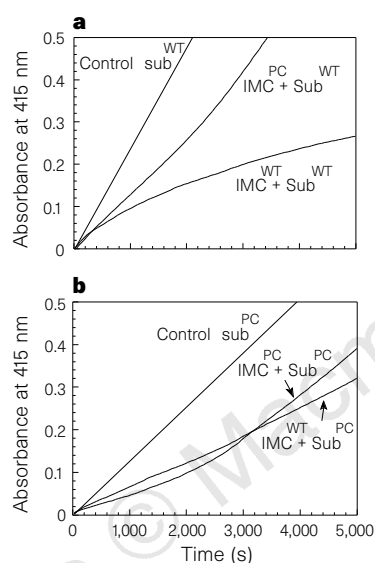
errata

Protein memory through altered folding mediated by intramolecular chaperones

U. P. Shinde, J. J. Liu & M. Inouye

Nature 389, 520–522 (1997)

Three of the superscripts in Fig. 3b of this Letter were incorrect. The amended figure is shown below. ☐



G-protein-coupled receptor of Kaposi's sarcoma-associated herpesvirus is a viral oncogene and angiogenesis activator

Carlos Bais, Bianca Santomasso, Omar Coso, Leandros Arvanitakis, Elizabeth Geras Raaka, J. Silvio Gutkind, Adam S. Asch, Ethel Cesarman, Marvin C. Gershengorn & Enrique A. Mesri

Nature 391, 86–89 (1998)

The name of the penultimate author of this Letter was misspelled: his name is Marvin C. Gershengorn. ☐

correction

A GPI-linked protein that interacts with Ret to form a candidate neurturin receptor

Robert D. Klein, Daniel Sherman, Wei-Hsien Ho, Donna Stone, Gregory L. Bennett, Barbara Moffat, Richard Vandlen, Laura Simmons, Qimin Gu, Jo-Anne Hongo, Brigitte Devaux, Kris Poulsen, Mark Armanini, Chika Nozaki, Naoya Asai, Audrey Goddard, Heidi Phillips, Chris E. Henderson, Masahide Takahashi & Arnon Rosenthal

Nature 387, 717–721 (1997)

An error in the designation of address symbols has caused confusion over the following authors' affiliations: R.D.K.'s present address is at Deltagen Inc.; L.S. Q.G. and A.G. are at Genentech Inc. ☐

KNOW YOUR COPY RIGHTS RESPECT OURS

The publication you are reading is protected by copyright law. Photocopying copyright material without permission is no different from stealing a magazine from a newsagent, only it doesn't seem like theft.

If you take photocopies from books, magazines and periodicals at work your employer should be licensed with CLA.

Make sure you are protected by a photocopying licence.



The Copyright Licensing Agency Limited
90 Tottenham Court Road, London W1P 0LP
Telephone: 0171 436 5931 Fax: 0171 436 3986